



Probabilistic combination of stellar astrophysical parameter estimates based on spectra, astrometry, photometry and the HR Diagram

prepared by: Coryn A.L. Bailer-Jones
Max Planck Institute for Astronomy, Heidelberg
Email: calj@mpia.de

approved by:

reference: GAIA-C8-TN-MPIA-CBJ-049

issue: 1

revision: 1

date: 2010-03-29

status: Issued

Abstract

Stellar effective temperature and line-of-sight extinction estimates from BP/RP spectra show strong degeneracies at lower signal-to-noise ratios. I develop a probabilistic framework in which we can use Gaia measurements of G and parallax together with our knowledge of stellar evolution from the Hertzsprung–Russell Diagram (HRD) to further constrain stellar parameters. This makes better use of all available information, enforces self-consistency and can easily be extended to include prior information and other data. Here I give both the theoretical development and a simple proof-of-concept demonstration.

Document History

Issue	Revision	Date	Author	Comment
1	1	2010-03-29	CBJ	Corrected an error in section 3.1, equation 15 (doesn't affect any results/demonstrations) and in statement of assumptions; Modified section 3.3 to discuss interpretation/use of the HRD as a prior; Comments from R. Drimmel.
1	0	2010-02-01	CBJ	Degeneracy map replaced with likelihood map; section 1.2 removed; section 2 improved; Demonstration and Discussion added. Minor changes throughout. Comments from P. Tsalmantza
D	3	2009-12-11	CBJ	Equation 16 corrected; HRD prior added
D	2	2009-12-07	CBJ	Corrected q PDF (and thus the equation for $P(A, T \mathbf{p}, q)$) and explained the parallax/mag constraint better in new section 1.1
D	1	2009-11-23	CBJ	Incomplete draft (HRD prior not parametrized; demonstration not done)

Contents

1	The problem	4
1.1	Parallax/magnitude constraint	6
2	Theory of probabilistic combination	7
3	Simple models	9
3.1	Parallax/magnitude noise model	9
3.2	Priors	10
3.2.1	Extinction in a band vs. the extinction law	12
3.3	Hertzsprung–Russell Diagram	13
4	Demonstration	14
5	Extension to other APs and additional data	21
5.1	Additional data	21
5.2	Additional APs	22
6	Discussion	23

TABLE 1: Notation

ϕ	vector of stellar astrophysical parameters (APs)
A	(= A_G) interstellar extinction in the G band (mag)
T	stellar effective temperature (K)
M	absolute magnitude in the G band (mag)
m	(= G) apparent magnitude in the G band (mag)
ϖ	parallax (arcsec)
b	Galactic latitude (degrees)
q	= $m + 5 \log \varpi$
\mathbf{p}	normalized BP/RP spectrum
\mathcal{M}	initial stellar mass (M_\odot)
τ	stellar age (Gyr)
log	base 10 logarithm
ln	natural logarithm

1 The problem

AP estimation algorithms usually estimate the stellar APs based only on the spectrum or colours. While some algorithms only provide the “best” estimates, others may provide a probability density function (PDF) over the APs, $P(\phi|\mathbf{p})$ (see Table 1 for my notation). Depending on the specific algorithm used, this may involve an implicit prior PDF over the APs (e.g. one influenced by the training data set; see Bailer-Jones et al. 2008). As has been shown in Bailer-Jones (2010a,b), the APs A (interstellar extinction) and T (effective temperature) are strongly degenerate with one another. The term “degenerate” is used here to mean “correlated over a significant part of the AP range(s) in a complex manner in a way that is not accurately described by the expected covariance”. Generally this means that $P(\phi|\mathbf{p})$ is multimodal or shows a broad ridge, both of which have been observed for APs estimated from BP/RP (Bailer-Jones 2010b).

The question arises whether we can reduce this AP degeneracy using additional information. In the context of Gaia we have at least three separate pieces of information:

1. The likelihood function $P(\mathbf{p}|A, T)$, which is computed using the ILIUM forward model (Bailer-Jones 2010b). The likelihood map introduced in that TN is just $P(\phi|\mathbf{p})$ assuming a uniform prior on ϕ , so is equal to the likelihood function to within a multiplicative factor.¹ I here neglect any dependence on other APs, e.g. the weak APs $\log g$ and $[\text{Fe}/\text{H}]$.
2. Gaia’s measurements of the source apparent magnitude and its parallax put some

¹Later on I assume that the likelihood map is a function of the extinction in the G band (A_G) rather than the extinction parameter (A_0). ILIUM was used in Bailer-Jones (2010a,b) to estimate the latter, but could be used to estimate the former. I do not expect the general results – in particular the extent of the degeneracy – to be significantly different in that case. However, this difference does have an impact on how we represent or interpret the priors on A , which will be discussed in section 3.2.1.

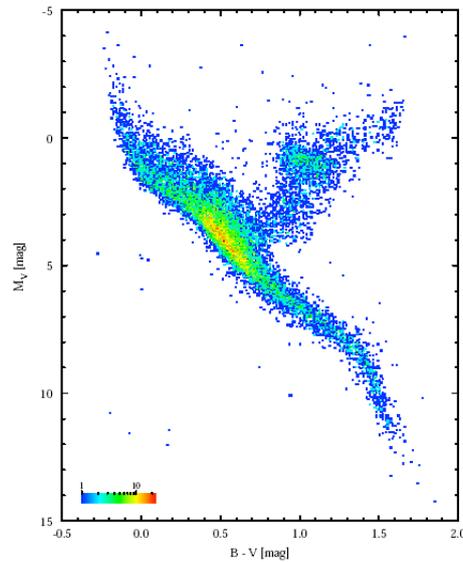


FIGURE 1: Hipparcos Hertzsprung–Russell Diagram

constraint on M (the absolute magnitude) and A (as explained in section 1.1).

3. Our knowledge of stellar structure and evolution – the HRD – likewise constrains M and T , because it is not uniformly populated (Fig. 1). If we ignore metallicity dependence we can just think of the HRD as the prior PDF $P(M, T)$, prior in the sense of being independent of a Gaia measurement of an individual star.²

By combining these pieces of information properly we can ensure astrophysical consistency between the derived parameters and the data. As an example, for a given observed magnitude and distance, there is a limit on the amount of extinction because stellar astrophysics doesn’t permit arbitrary values of the intrinsic luminosity. Here I show how these three pieces of information may be combined probabilistically in order to improve the estimates of the stellar APs beyond use of the spectrum alone, and to ensure consistency. Inclusion of other pieces of information, such as RVS or AP estimates from external catalogues, I will consider later in section 5.

By “spectrum” I mean here the *relative* variation in flux (or photon counts) with wavelength. That is, the spectrum has been normalized in some way to remove the total flux information. We can achieve this in at least two different ways: either area normalization (dividing the spectrum by its integral over wavelength) or by G-band normalization (dividing the spectrum by $10^{-0.4G}$). In the first case the spectrum has no absolute flux information so is strictly independent of G . In the latter case, because the profile of G is not quite the same as the combined profile of BP and RP, the absolute flux removal is imperfect so the spectrum becomes only approximately independent of G .

²There is no logical difference between a “prior”, a “likelihood” and a “posterior”. They are all probabilities, differing only in what data and assumptions they are conditioned on.

1.1 Parallax/magnitude constraint

Point (2) in the above list can be seen from simple geometry and the definition of absolute magnitude

$$m + 5 \log \varpi = M + A - 5 . \quad (1)$$

(In Gaia notation $m = G$ and $M = M_G$, but in the theoretical part of this TN I use this more general notation.) Clearly the parallax and magnitude constrain M and A , yet the former are not known exactly due to noise. So how can we use this? We need to define a noise model. For brevity I define

$$q = m + 5 \log \varpi \quad (2)$$

which is our noisy Gaia measurement; equation 1 only holds in the absence of noise. So more generally we consider the quantity

$$x = q - (M + A - 5) . \quad (3)$$

The noise model for x is $P(x|M, A)$, which has expectation value $E[x] = 0$ and variance σ_q^2 , the variance in q (as M and A are not measured). Let us model this as a 1D Gaussian in x , $N_x(0, \sigma_q)$. For a given star (fixed M and A), $P(x|M, A)$ has its maximum when the measurement q equals $M + A - 5$. The further a measurement of q (and thus x) is away from this value the less probable it is. Note that as q is the only noisy term in equation 3 it follows that $P(x|M, A) = P(q|M, A)$. Proof:

$$\begin{aligned} P(x|M, A) dx &= P(q|M, A) dq \\ P(x|M, A) &= P(q|M, A) \frac{dq}{dx} \\ &= P(q|M, A) \times 1 . \end{aligned} \quad (4)$$

Now consider $P(q|M, A)$ as a function of M and A for a given measurement q . This is shown in Fig. 2. We can think of proposing trail solutions for M and A : the further they lie from the solid line, the lower $P(q|M, A)$ (inset in Fig. 2). How quickly the probability drops off as we move away from this line depends on how accurate we think q is, which is set by σ_q . Hence with the Gaussian approximation of the noise model for q we have

$$P(q|M, A) = N_x[0, \sigma_q(m, \varpi)] = N_x[q - (M + A - 5), \sigma_q(m, \varpi)] \quad (5)$$

the transformation of the mean following from equation 3. This is a 1D Gaussian over the variable x with mean and standard deviation as given, or equivalently, a Gaussian over the variable $(M + A - 5)$ with mean q . As expected from equation 1, a measurement of m and ϖ constrains the possible values of M and A . Equation 5 quantifies this using the known statistics of the noise of the Gaia measurements. Note that this does not constrain M or A to have astrophysically “sensible” values (e.g. the line continues to negative A in Fig. 2). This may be done by the HRD prior and/or a prior on extinction.

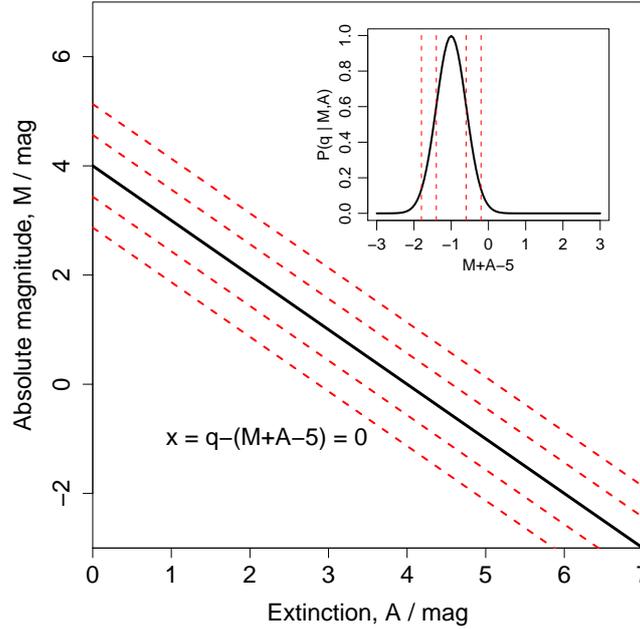


FIGURE 2: Illustration of using the parallax and apparent magnitude ($q = m + 5 \log \varpi$) to constrain extinction and absolute magnitude. Here we measure $q = -1$, which corresponds to a $m = 14$ star at 1 kpc, for example (or to a $m = 19$ star at 10 kpc, etc.). If this were a noise-free measurement, it would constrain the solution $\{M, A\}$ to lie only on the solid black line. But as q is a noisy measurement – taken here to be 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 show in the inset panel, the red dotted lines in both plots showing the 1 and 2 sigma levels for this value of σ_q (which is characteristic for stars around $G=19$ at a distance of kpc or so).

2 Theory of probabilistic combination

Here I derive an expression for $P(A, T | q, \mathbf{p})$ based on the three pieces of information listed in section 1. The discussion of conditional independence in Bailer-Jones (2010b) may be useful background material.

Bayes' theorem tells us that

$$P(A, T | \mathbf{p}, q) = \frac{P(\mathbf{p}, q | A, T) P(A, T)}{P(\mathbf{p}, q)} \quad (6)$$

and from the axioms of probability (Cox 1946) we have

$$P(\mathbf{p}, q | A, T) = P(\mathbf{p} | q, A, T) P(q | A, T) . \quad (7)$$

As \mathbf{p} and q are independent measurements (conditioned on A and T)

$$P(\mathbf{p} | q, A, T) = P(\mathbf{p} | A, T) . \quad (8)$$

Using this in equation 7 and substituting that into equation 6 gives

$$\begin{aligned}
 P(A, T | \mathbf{p}, q) &= \frac{P(\mathbf{p} | A, T) P(q | A, T) P(A, T)}{P(\mathbf{p}, q)} \\
 &= \frac{P(\mathbf{p} | A, T) P(q | A, T) P(A) P(T)}{P(\mathbf{p}) P(q)} \quad (9)
 \end{aligned}$$

where the second line follows if we make the reasonable assumptions that both (1) \mathbf{p} and q and (2) A and T , are unconditionally independent. The terms $P(A)$, $P(T)$, $P(\mathbf{p})$ and $P(q)$ are priors over these APs or measurements. The first term in the numerator is equal to the likelihood map (to within a normalization constant). The second term we need to further decompose, plus we want to introduce some dependence on M so we can incorporate the HRD and equation 5. A general rule of probability is $P(z) = \int P(z, y) dy = \int P(z | y) P(y) dy$ where the integral is over all y . Thus we can write

$$P(q | A, T) = \int P(q | M, A, T) P(M | A, T) dM . \quad (10)$$

It is clear from equation 2 that q is determined entirely by M and A (plus uncorrelated noise). So once q is conditioned on these two parameters, T adds no further information. Therefore the first term in equation 10 is $P(q | M, A, T) = P(q | M, A)$. (Note that this does *not* require that T be independent of M and/or A .) The second term in the integral we can rewrite using Bayes' theorem

$$\begin{aligned}
 P(M | A, T) &= P(A, T | M) \frac{P(M)}{P(A, T)} \\
 &= P(A | T, M) P(T | M) \frac{P(M)}{P(A, T)} \quad \text{from general property of joint probabilities} \\
 &= P(A) P(T | M) \frac{P(M)}{P(A) P(T)} \quad \text{as } A \text{ is independent of } M \text{ and } T \\
 &= P(T | M) \frac{P(M)}{P(T)} \\
 &= P(M | T) \quad \text{from Bayes' theorem} \\
 &= \frac{P(M, T)}{P(T)} . \quad (11)
 \end{aligned}$$

This result should be obvious because given T , A tells us nothing additional about M .³ Substituting equation 11 into equation 10 and using the simplification for its first term described above gives

$$P(q | A, T) = \int P(q | M, A) \frac{P(M, T)}{P(T)} dM . \quad (12)$$

³ M and T are properties of the star only, whereas A is not: unconditioned on any measurement of the spectrum or parallax, knowledge of M and/or T does not constrain A . Note, in contrast, that M and T are not independent of each other due to the HRD.

We now substitute this into equation 9 to give the final result

$$P(A, T|\mathbf{p}, q) = \underbrace{P(\mathbf{p}|A, T)}_{\text{likelihood}} \underbrace{\frac{P(A)}{P(\mathbf{p})P(q)}}_{\text{priors}} \underbrace{\int \underbrace{P(q|M, A)}_{q \text{ constraint}} \underbrace{P(M, T)}_{\text{HRD map}} dM}_{\text{HRD/q factor}} \quad (13)$$

where we see that $P(T)$ has cancelled. The result is a 2D PDF over A and T given the measured spectrum, \mathbf{p} , and the measured parallax/photometry, q . This can be seen as a product of three terms. The first term is the likelihood function (equal to the likelihood map to within a constant factor), the second term contains priors over the extinction and the data. The third term is an integral over two factors: the combined astrometric/photometric noise model and the HRD. The integral marginalizes over the unknown M to give what I will call the “HRD/q factor” and is a function of A and T .

If we lack information then some terms in equation 13 simplify. For example, if we have no measurement of q then we can set the q constraint to a constant. If we don’t want to use an informative prior on the extinction we can set $P(A)$ to a constant. Likewise, if we don’t want to use the HRD prior, then this is equivalent to setting $P(M, T)$ to a flat distribution (!), in which case the second term under the integral is also a constant.

Once we have evaluated $P(A, T|\mathbf{p}, q)$, we can separately marginalize over A and T in order to get final estimates for both APs, i.e. $P(T|\mathbf{p}, q) = \int P(A, T|\mathbf{p}, q)dA$, and likewise for A .

Once we have written down specific models for the terms in equation 13 we can use some sampling algorithm to calculate it. If we take the logarithm we get the log posterior equal to a sum of terms, each depending on different information. In maximizing this posterior with respect to A and T the relative sizes of the terms will change. Thus we can consider the AP inference problem as an optimization problem according to various “data”, “constraints” or “priors”.

3 Simple models

To apply the theory we need to write down specific models for the terms in equation 13.

3.1 Parallax/magnitude noise model

I have assumed that the noise model in q can be approximated at a 1D Gaussian (equation 5)

$$P(q|M, A) = N[q - (M + A - 5), \sigma_q(m, \varpi)] \quad (14)$$

If the standard deviations of the photometric noise and the parallax noise are σ_m and σ_ϖ respectively (this does not assume their individual distributions to be Gaussian), then from the basic

relationship $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$ we have

$$\sigma_q(m, \varpi)^2 = \left(\frac{5}{\ln 10} \frac{\sigma_\varpi}{\varpi} \right)^2 + \sigma_m^2 + 2 \left(\frac{5}{\ln 10} \frac{\sigma_\varpi}{\varpi} \right) \sigma_m \rho(m, \log \varpi). \quad (15)$$

The covariance between m and $5 \log \varpi$ has been written in terms of their correlation coefficient, ρ , where $\text{Cov}(m, 5 \log \varpi) = \rho(m, \log \varpi) \sigma_m \sigma_{5 \log \varpi}$. Note that $\rho(m, 5 \log \varpi) = \rho(m, \log \varpi)$.

For a source at $G = 18.5$ the parallax accuracy is about $120 \mu\text{as}$ (de Bruijne 2009)⁴, i.e. $\sigma_\varpi = 1.2e^{-4}$. The photometric accuracy is presumably limited by calibration errors, which I take to be 3 mmag ($\sigma_m = 0.003$) and therefore uncorrelated with the astrometric accuracy ($\rho(m, \varpi) = 0$). Thus for a star at a distance of 1 kpc, equation 15 gives $\sigma_q = 0.26$ mag (the astrometric error makes up essentially all of this). At 100 pc and 5 kpc we get $\sigma_q = 0.026$ mag and 1.30 mag respectively. If we adopt a $300 \mu\text{as}$ parallax accuracy at $G = 20$, then for a star of this magnitude at 1 kpc, 5 kpc and 20 kpc we get $\sigma_q = 0.65$ mag, 3.26 mag and 13.0 mag respectively. (Note that because the astrometric error dominates, σ_q scales linearly with distance at a fixed apparent magnitude.)

3.2 Priors

One cannot dispute the *existence* of priors (information which is independent of the specific measurements considered) and in practice we always use them, even if implicitly. In the weakest form they are constraints, e.g. T and A cannot be negative. Yet we have prior knowledge beyond that, in the sense that extreme values (e.g. 100 magnitudes of extinction or an effective temperature of 125 K) are very unlikely. Priors put this on a rigorous basis. What one can and should consider carefully is *what* priors to use.

Equation 13 contains explicit priors in \mathbf{p} , q and A . The first two we consider as “direct” measurements. It is important to realise that these two priors are *unconditional on the APs*, so the reasonable prejudice (prior knowledge) that hot stars are rarer than cool ones is not relevant in assigning distributions to these. Moreover, as we are interested in inferring the probabilities of solutions in the (A, T) plane for a *given* measurement (\mathbf{p}, q) , the prior over the data we can calculate as the normalization constant

$$P(\mathbf{p}, q) = \int_{A, T} P(\mathbf{p}|A, T) P(A) \left(\int P(q|M, A) P(M, T) dM \right) dA dT. \quad (16)$$

and isn’t really interesting as it doesn’t distinguish between different solutions of A and T . In practice most of the terms on the right-hand-side of equation 13 are not individually normalized; to get actual probabilities we simply perform this global normalization of $P(A, T|\mathbf{p}, q)$ at the end.

The prior on A , on the other hand, is relevant. We could assign a non-informative prior (and in fact I use this in most of the demonstrations), but we should do better than this, because

⁴This is after the nominal five years of observations; it varies with sky position (ecliptic latitude) due to the Gaia scanning law.

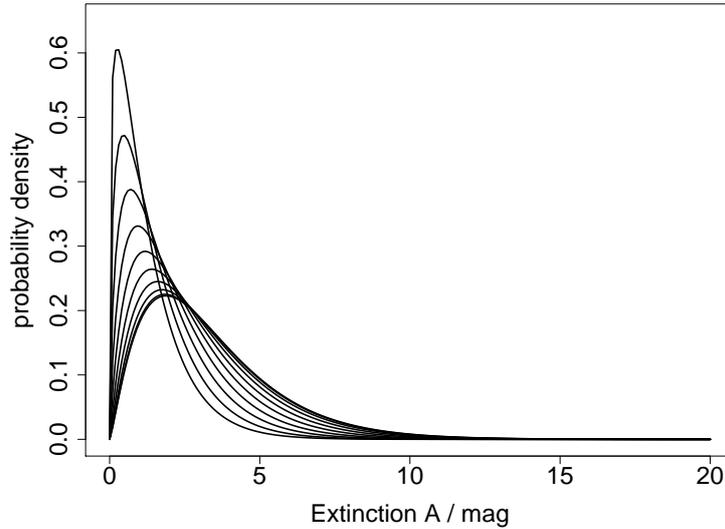


FIGURE 3: Possible prior PDF over extinction using the gamma function. The curves with peaks from left to right are for Galactic latitude decreasing from 90° to 0° in steps of 10° .

our reasonable prior expectations of A will help to better determine T even in the absence of measuring the parallax or photometry. Let us try and produce a simple prior.

First, we have the constraint that $A \geq 0$. For a magnitude-limited survey ($m = 20$) there is also an upper limit on A for objects which we can see. A very extreme case would be a magnitude $M = -5$ giant star at 100 pc observed at $m = 20$, which implies it has $A = 20$ (equation 2). I will take this an upper limit. (The maximum extinction used in the CU8 GOG simulations up to and including cycle 8 is $A_0 = 10$ mag.) We could adopt a prior which is uniform between these two values, and zero beyond, but the discontinuity at $A = 20$ should be avoided.

The line-of-sight extinction varies considerably over the sky, primarily as a function of Galactic latitude, b . As object coordinates are invariably known, it is sensible to use a simple model of $A(b)$.⁵ Let's write down a model which gives a broader distribution of extinctions at lower latitudes than at higher ones. A simple smooth model is the gamma function in which both the shape parameter k and the scale parameter θ are given a dependence on latitude. Specifically we could consider

$$P(A|b) = A^{k-1} \frac{e^{-A/\theta}}{\Gamma(k)\theta^k} \quad (17)$$

where $k = \cos(b) + \frac{5}{4}$ and $\theta = \frac{1}{2} \cos(b) + 1$. This is plotted for several values of b in Fig. 3. This is not a particularly good model. For example the extinction range should drop more quickly with latitude (e.g. Marshall et al. 2006). On the other hand, we might not want to assign such

⁵We should avoid introducing too much current knowledge of Galactic structure into estimating APs from Gaia data, as Gaia is intended to better determine that structure. On the other hand, no one seriously believes Gaia will dispose of the concept of the Galactic disk, so a simple model is warranted.

very low probabilities to high extinctions because of the possibility of seeing objects in a star forming region or cloud. (If such clouds could be identified based on source clustering then this information may motivate us to use an alternative prior in those cases). One can – and should – debate how “strong” a prior we use, but it does not change the principle. We’ll see in section 4 that this is actually quite a weak prior compared to the extinction constraints coming from other data.

There are suitable alternative distributions. If we wanted a prior with a maximum at $A=0$ (and zero gradient there) which still asymptotes to zero at infinite extinction, then the Gaussian with zero mean is useful. If desired, a closer match to the Galaxy would require some nonparametric smoothing of observations. If we wanted to use a more complex model depending also on distance, $P(A|b, \varpi)$, then we would have to develop the theoretical part differently from equation 9 on in order to include this term. Not having any distance dependence is at least conservative (i.e. A and T not over-constrained by a more informative prior).

3.2.1 Extinction in a band vs. the extinction law

In the above I have put a prior on A , the extinction in the G band for a given star, and not on the parameter A_0 of the extinction law for the line-of-sight to that star. This law is

$$A_\lambda = A_0[a(\lambda) + b(\lambda)/R_0] \quad (18)$$

where A_λ is the extinction in magnitudes at wavelength λ , $a(\lambda)$ and $b(\lambda)$ are polynomials and A_0 and R_0 are parameters fitted from data (Cardelli et al. 1989).⁶ When we consider an extinction map, we usually consider the variation of A_0 across the sky, because this is a property only of the interstellar medium (as is R_0). In contrast, A is the integral of this function over the spectrum, $F_\lambda(\phi)$, and the filter pass band, g_λ ,

$$A = -2.5 \log \left(\frac{\int F_\lambda(\phi) g_\lambda 10^{-0.4A_\lambda} d\lambda}{\int F_\lambda(\phi) g_\lambda d\lambda} \right) \quad (19)$$

and therefore depends on the specific star. My latitude-dependent prior above is actually not self-consistent because it assumes that no matter what the type of star, it has the same $P(A)$ (for a given line-of-sight) and therefore a $P(A_0)$ which depends on the type of star. It would be more natural to have a common probability for A_0 , which would correspond to a range of probabilities for A depending on the type of star. As the theory – in particular the q constraint – works with A rather than A_0 , we would have to calculate A (and the corresponding prior distribution over this) from A_0 via equation 19. Yet this depends on the other APs, in particular T_{eff} , which is degenerate with A_0 . (Getting a likelihood map over A_0 rather than A is not a problem.) We need not enter this complexity now: the use of a prior over A rather than A_0 is acceptable provided we only want a weak prior.⁷

⁶CU8 now uses A_0 and R_0 as labels for these fitting constants, rather than the traditional A_V and R_V , in order to emphasise that A_V is not the extinction in the V band. However, in order to be consistent with previous ILIUM technical notes, in section 4 I often use A_V as a synonym for A_0 .

⁷We could alternatively change the theory to work with A_0 rather than A_G , but it’s really the latter we want to know. The extinction parameter assumes the extinction law, whereas not all Galactic extinction follows this. An

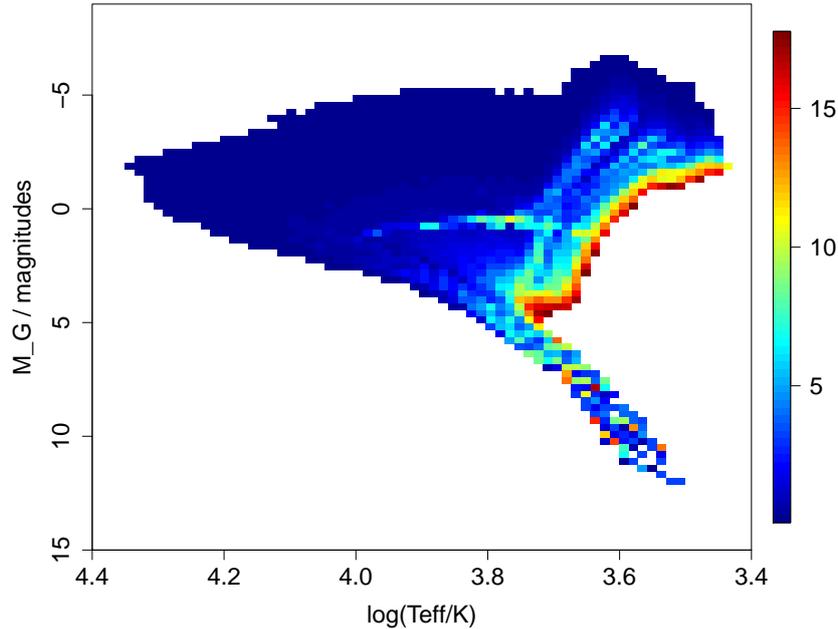


FIGURE 4: HRD map (unnormalized $P(M, T)$) based on the Girardi et al. (2000) isochrones interpolated to a regular grid. The colour of each cell in the grid indicates the average age (in Gyr) of the stars in that cell.

3.3 Hertzsprung–Russell Diagram

The HRD, $P(M, T)$ is also a prior, but one motivated by well-established astrophysics. Years of astrophysical research have shown us that the HRD is not uniformly populated, with significant empty regions and significant variations in density (and therefore prior probability density). Even a crude use of the HRD should allow us to down-weight some solutions to (M, T) (although whether this also constrains (A, T) in practice will be evaluated below).

If (1) star formation proceeded at a constant rate, (2) stars of all masses were equally common, and (3) all stars could be equally well observed (no practical detection limit), then the probability of observing a star at any one point in the HRD is proportional to the amount of time it spends there. This duration at any point is a difficult piece of information to extract from published theoretical isochrones (at least it is for the ones used here), I make the additional rather simplistic assumption that the speed of evolution across any point of the HRD is inversely proportional to the star’s age. In that case the duration (and thus the probability of observing a star) is simply proportional to its age, τ .⁸ This is not at all astrophysically accurate, but it will do for

example is gray extinction, which has $E(B - V) = 0$ implying R_0 is infinite (Gorbikov & Brosch 2010). On the other hand we must adopt an extinction law in order to simulate the affects of extinction as needed by the spectral AP estimation algorithm.

⁸If s denotes the position in the HRD and $v(s)$ is the speed of evolution between two points s_1 and s_2 , taken to define a “cell” in the HRD, then the time spent moving across this cell is $\Delta t = \int_{s_1}^{s_2} \frac{1}{v(s)} ds = \Delta s / v(s)$, assuming $v(s)$ to be constant over the cell. For a regular grid of cells, Δs is the same over the whole HRD. The simplistic

the sake of a simple demonstration.

I construct such a probability map using the isochrones of Girardi et al. (2000). This provides the age, $\log(T_{\text{eff}})$ and absolute M_V and M_I band magnitudes (among other things) for 64381 stars with a range of masses (0.15 to $6.6 M_{\odot}$) and metallicities ($Z = 0.0004$ to 0.03). To achieve a regular grid I interpolate the Girardi et al. data onto a regular grid and average the ages for all stars which fall into a particular cell. I convert the M_V magnitude to Gaia M_G using the transformation of Jordi (2009) for unreddened stars (her Table 4)

$$M_G = M_V - 0.0244 - 0.0513(M_V - M_I) - 0.1762(M_V - M_I)^2 + 0.0097(M_V - M_I)^3. \quad (20)$$

The result is shown in Fig. 4. The regions which are white contain no stars, and so strictly exclude the possibility of any solution (prior probability of zero). In general we would not want to be so extreme and would replace zero with a very low probability. We may also want to smooth the prior (e.g. with kernel density estimation) rather than use a discrete grid. Note also that this model excludes white dwarfs, brown dwarfs, binaries and massive stars.

The simplistic age/duration assumption aside, I have also assumed that stars of different initial masses are equally common (a flat initial mass function, IMF). A more realistic IMF may be incorporated by multiplying the average age, τ , in a cell with the average initial stellar mass, \mathcal{M} , in that cell. This propagates the IMF through the whole HRD. A prior constructed in this way based using the Salpeter IMF, $dN \propto \mathcal{M}^{-2.35} d\mathcal{M}$, is shown in Fig. 5. This is plotted on a log colour scale as now the massive stars are much rarer.

Notwithstanding the deficiencies of this prior, it is sufficient to enable us to demonstrate the principle of the probabilistic combination method.

4 Demonstration

Let us now see to what extent equation 13 and the models presented in section 3 for the terms in that equation can reduce the degeneracies in the estimates of A and T for stars at a range of A and T . In all cases I use the likelihood map for spectra at $G=20$ from Bailer-Jones (2010b), reproduced in Fig. 6. In practice we would use ILIUM (or other AP estimation algorithm) to find the best single set of APs for a star, and then calculate the likelihood map according to the observed spectrum. Here, however, I do not use the ILIUM search method at all. I simply assume that we have found a set of APs (corresponding to the red cross) and have calculated the likelihood map according to its expected spectrum.

For a $G=20$ star at a distance of 1 kpc, $q = 5$ mag and $\sigma_q = 0.65$ mag, which specifies $P(q|M, A)$. Combining this with the age-only HRD prior (Fig. 4) and a flat prior over A , equation 13 gives the posterior PDF shown in Fig. 7. Comparing with the likelihood map (which assumes all other terms in equation 13 are constant) we see significant differences. In particular, this posterior has removed much of the solution space at high T_{eff}/A_V . As we have used a flat extinction prior,

assumption is v_s is inversely proportional to the age. Hence $\Delta t \propto \tau$.

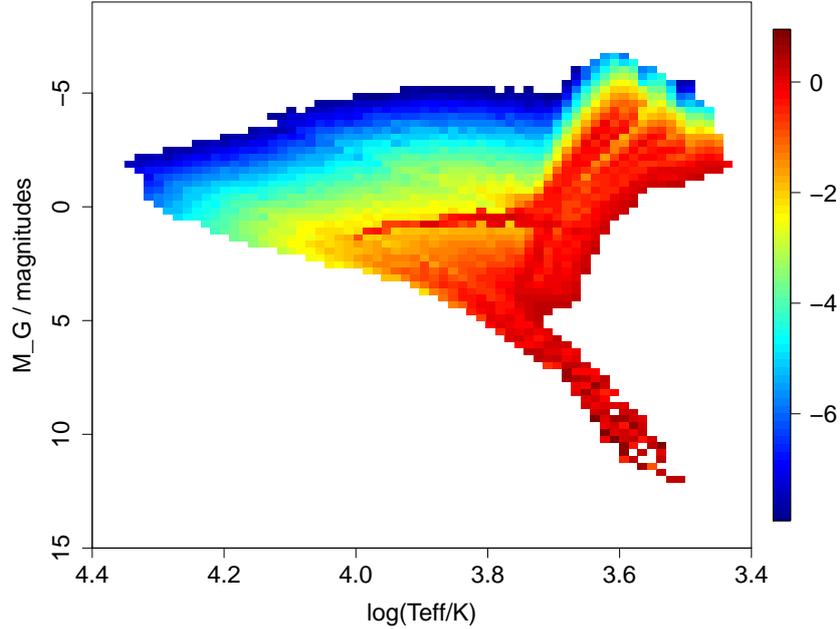


FIGURE 5: HRD map (unnormalized $P(M, T)$) based on the Girardi et al. (2000) incorporating the Salpeter IMF. The colour of each cell is proportional to the logarithm of the product of the average mass (in M_{\odot}) and age (in Gyr) in that cell.

this must be a result of the HRD/q factor (HRD map and q constraint in equation 13). We can see this if we plot the posterior PDF just for this term, i.e. without the likelihood function (we just set it to a constant in the calculation and then renormalize). This is shown in Fig. 8. (It’s independent of the A, T of the specific star, so it’s the same in all panels.) We can see how this term is excluding regions of the AP space.

How does the HRD/q factor achieve this? If we consider zero extinction, then $q = 5$ mag implies that $M = 5$ mag (or rather that it has a Gaussian distribution with this mean and a standard deviation of σ_q). The HRD map (Fig. 4) shows us that this limits considerably the possible values of T_{eff} ; a consequence of stellar physics. If A were higher, M would be lower, and a wider range of T_{eff} values is permitted by the HRD. Correspondingly we see the likely solution space for T_{eff} in Fig. 8 expanding as A increases. The “dent” in the solution space at intermediate A corresponds to the lack of physical stellar solutions in the HRD to the right of (cooler) and below (fainter) the giant branch.

We now see how the HRD/q factor combines with the likelihood map to produce the posterior PDF. The actual influence of course depends on the values of q and σ_q . If we now imagine the astrometry placed the star at 5 kpc instead of 1 kpc, what differences would be expected? As the apparent magnitude of the star is unchanged, we now expect solutions corresponding to intrinsically brighter stars to be more probable. As luminosity $\propto T^4$ (Stefan’s law) this translates to a higher probability for larger T_{eff} . At 5 kpc and $G=20$, $q = 1.505$ mag and $\sigma_q = 3.26$ mag. As q is smaller by 3.5 mag, we indeed see from the HRD that a wider T_{eff} range is permitted; the

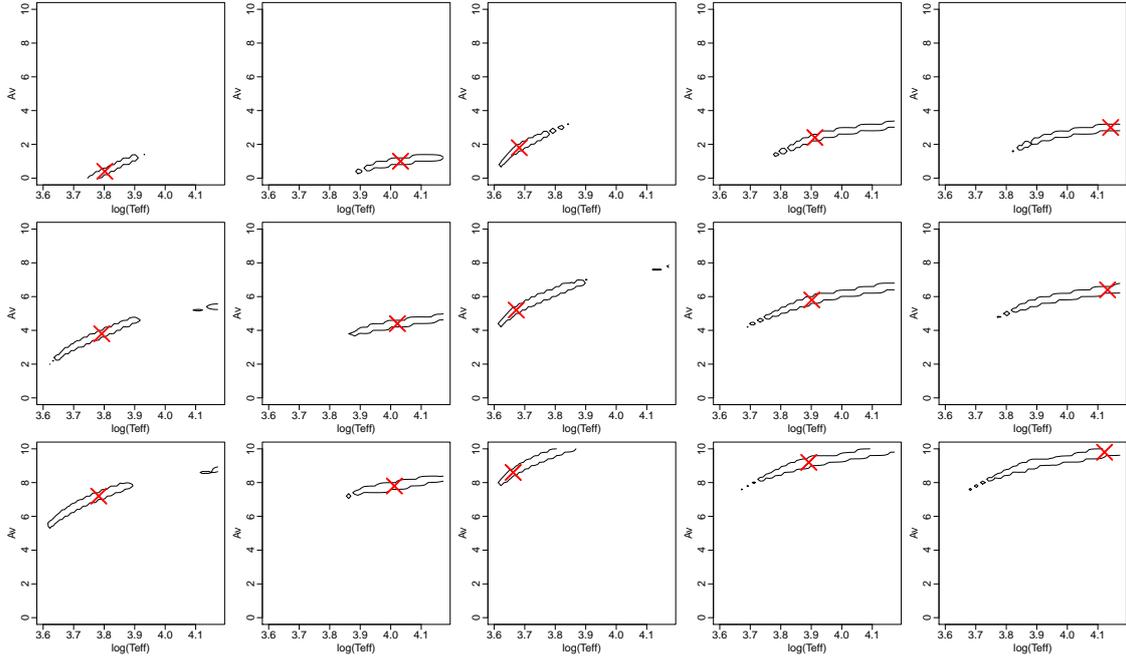


FIGURE 6: Likelihood map for a selection of stars at $G=20$. The red cross is the most probable solution and the contour contains the 99% most probable AP solutions

HRD/ q factor is shown in Fig. 10 and agrees with this expectation.

The corresponding posterior PDF including this HRD/ q factor with the likelihood function is shown in Fig. 9. It is informative to contrast this with the posterior PDF for the same spectra but observed at 1 kpc in Fig. 7. The differences are entirely consistent with the differences between their corresponding HRD/ q factor maps. A little consideration will lead one to realise that larger values of q tend to place stronger constraints on the posterior PDF. (Hint: larger q implies not only nearer but also more accurate parallaxes.)

In comparing the likelihood map with the posterior PDF at 1 kpc, and noting that the red cross lies outside of the 99% confidence region in several panels in the latter, we might be misled to conclude that the method is somehow giving “worse” solutions. However, the red cross is only the “correct” under the hypothesis that all AP solutions are a priori equally probable. It is only the most probable solution given the spectrum (only) under this same hypothesis. Astrophysically this is a rather unrealistic hypothesis. The HRD introduces the reality of stellar physics, and the q constraint adds in further data. The point is that we are not just classifying a spectrum, but rather trying to infer the APs of the star based on several pieces of information, of which the spectrum is just one. The stars corresponding to the 15 A, T combinations in the panels were not chosen from a realistic Galaxy model, and some actually correspond to very unlikely stars once we also specify G and ϖ . It is precisely this information which the HRD/ q factor introduces into our posterior PDF. Depending on the values of q and σ_q , the HRD may actually provide a very strong “prior” on A and T . This is what we get when we demand

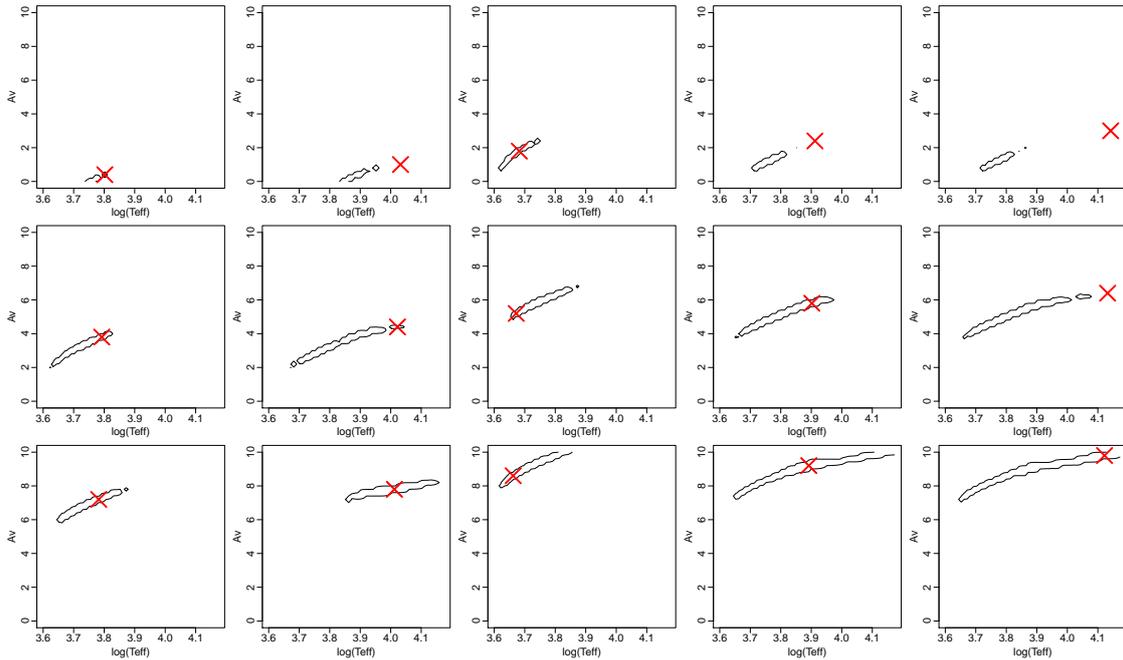


FIGURE 7: Posterior probability density function for G=20 spectra at 1 kpc ($q=5$ mag and $\sigma_q=0.65$ mag) for the age-only HRD prior and flat A prior and likelihood function shown in Fig. 6. The contour contains the 99% most probable AP solutions.

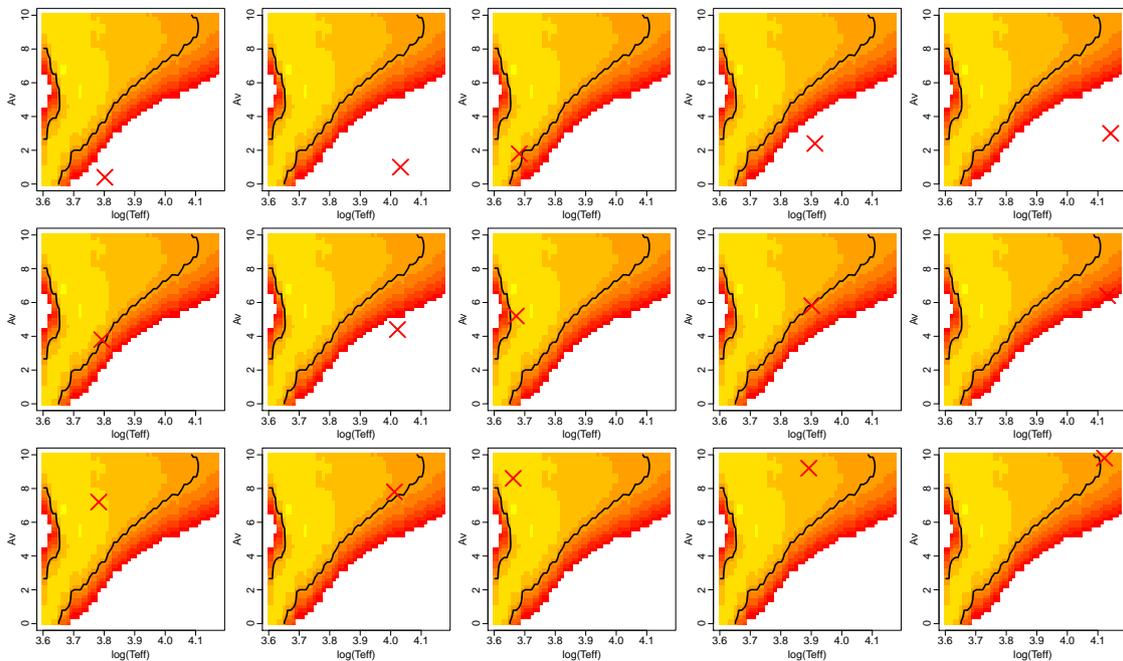


FIGURE 8: As Fig. 7 but only including the HRD/ q factor (age-only HRD prior), which is the same for all stars. To better show the surface I also plot the logarithm of the posterior as a colour scale, with higher probabilities brighter/yellow.

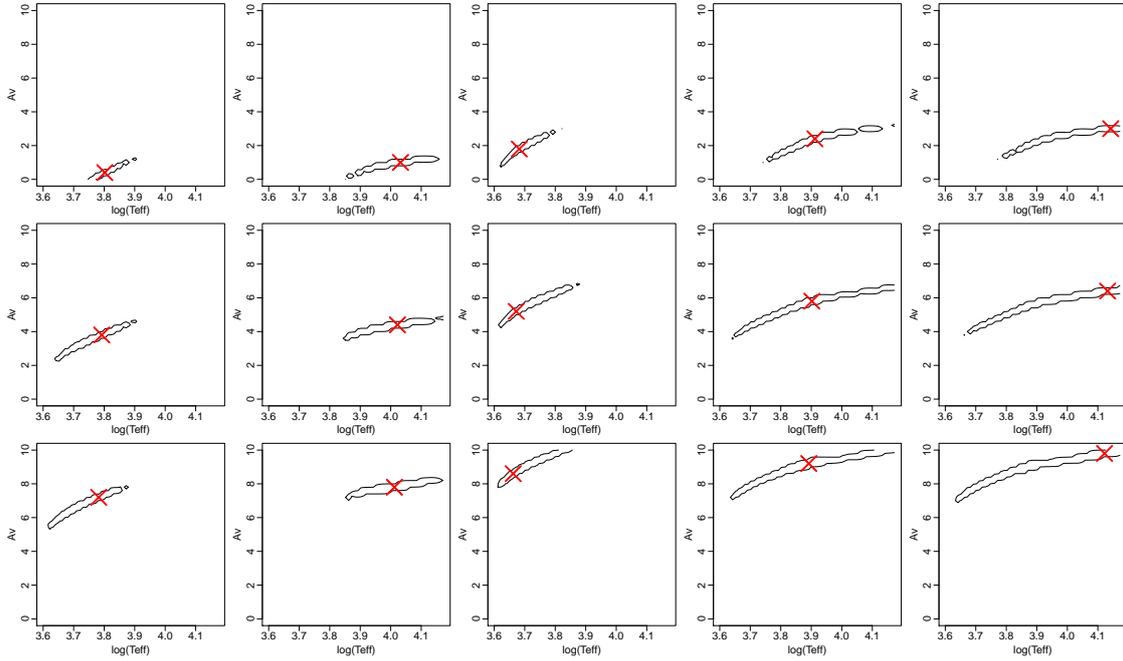


FIGURE 9: As Fig. 7 but for a star at 5 kpc and therefore for $q = 1.505$ mag and $\sigma_q = 3.26$ mag.

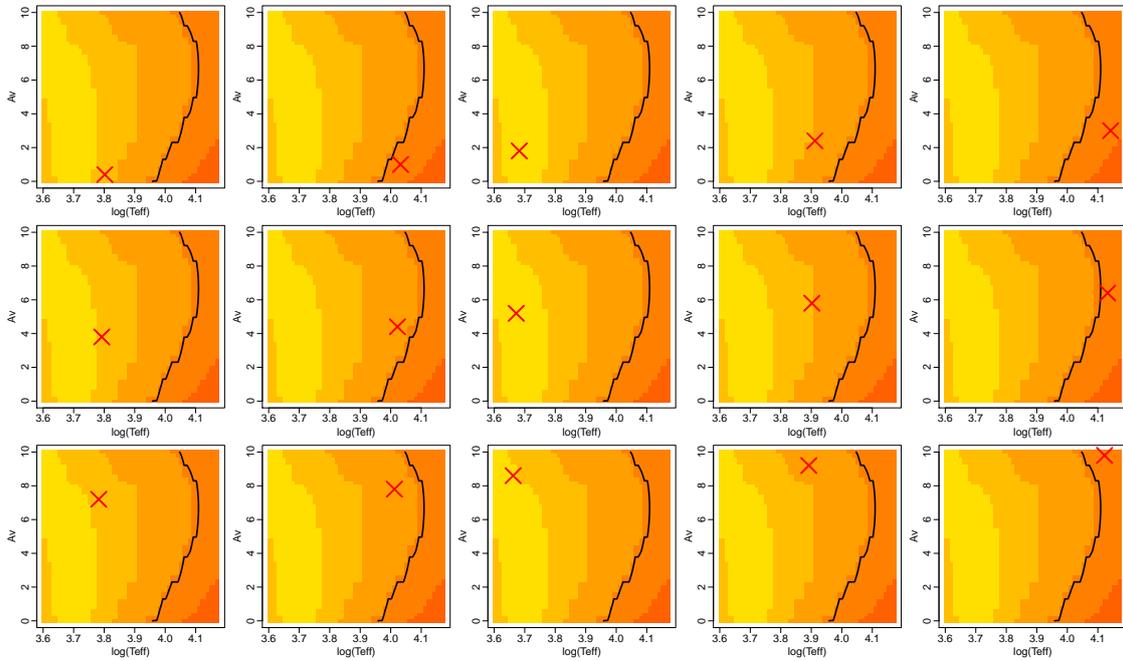


FIGURE 10: As Fig. 9 but only including the HRD/ q factor. (Higher probabilities are brighter / to the left of the 99% contour.)

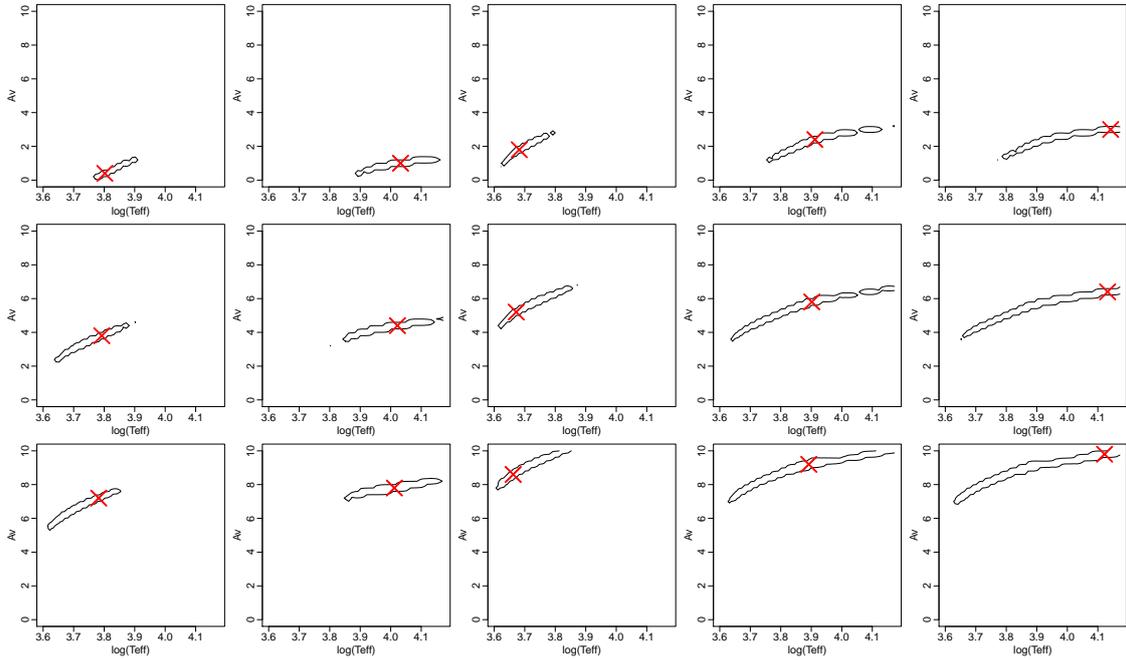


FIGURE 11: As Fig. 9 but now including the extinction prior in equation 17 for $b = 0$.

astrophysical self-consistency. Note that none of this calculation uses a Galaxy model. The q constraint rises entirely from geometry (flux falls off with the square of distance) and the definition of extinction, and the HRD is based only on stellar physics. A is affected by these because it is not independent when conditioned on these information and the spectrum.

The model in equation 13 is quite flexible. If we didn't have parallaxes we can set the q constraint to a constant. The HRD/ q factor – which is just a marginalization of the HRD over M – has no A dependence and so results in a term $P(T)$, a prior over T_{eff} only. In contrast, if we had the q constraint but set the HRD to be uniform this has no effect, as we can see from Fig. 2: For a given q at fixed A we get a certain distribution over M , which when integrated over M gives a certain value. However, this value will be the same for all A , as the distribution over M is independent of A . In that case the HRD/ q factor reduces to a constant.

What happens if we replace with age-only prior with the age-IMF combined prior of Fig. 5? Answer: virtually nothing – the posterior PDFs are hardly changed – for these stars and values of q . The reason is that the highest T_{eff} red cross is $\log(T_{\text{eff}}) = 4.14$ (13800 K), whereas the major differences between the two HRDs occur at higher temperatures.

All of the examples so far have assumed a flat prior over extinction. It is interesting – but now understandable – that the “loss” of solutions for high T_{eff}/A_V has been the result of demanding self-consistency between parallax and apparent magnitude on the one hand and stellar physics (the HRD) on the other. This has *not* been a consequence of an extinction prior!

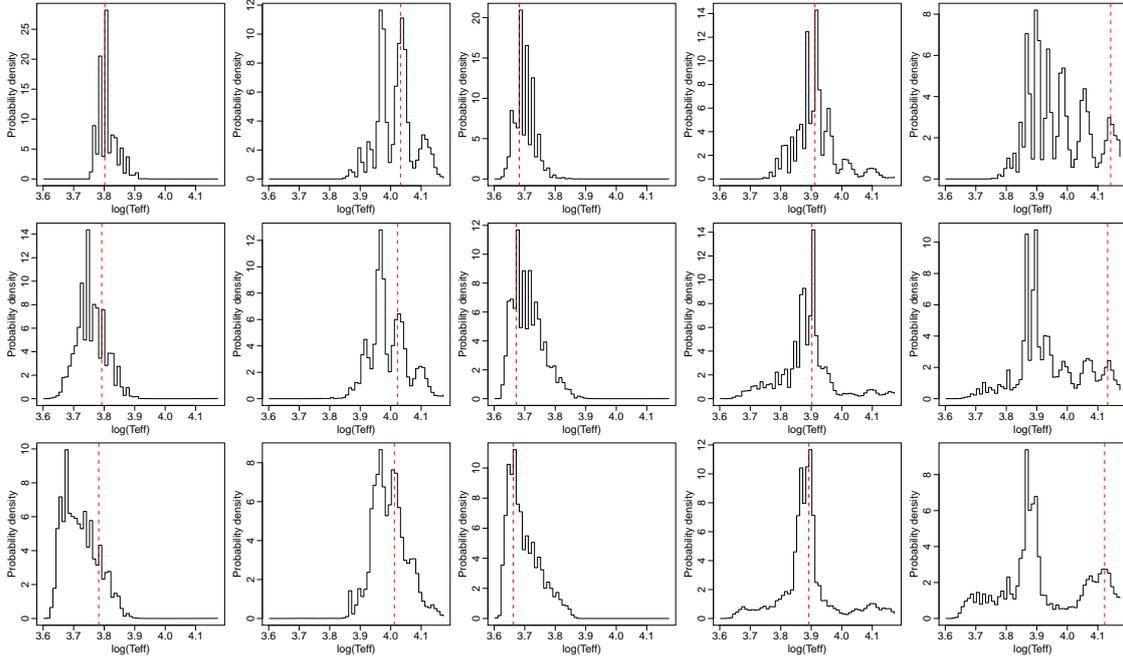


FIGURE 12: Posterior PDF over T_{eff} formed by marginalizing over A_V in Fig. 9 ($G=20$, 5 kpc, flat A_V prior, age-only HRD prior). The red dashed line shows the “true” APs according to just the spectrum. A deviation of the mean of the PDF from this does not indicate low performance, as discussed in the text.

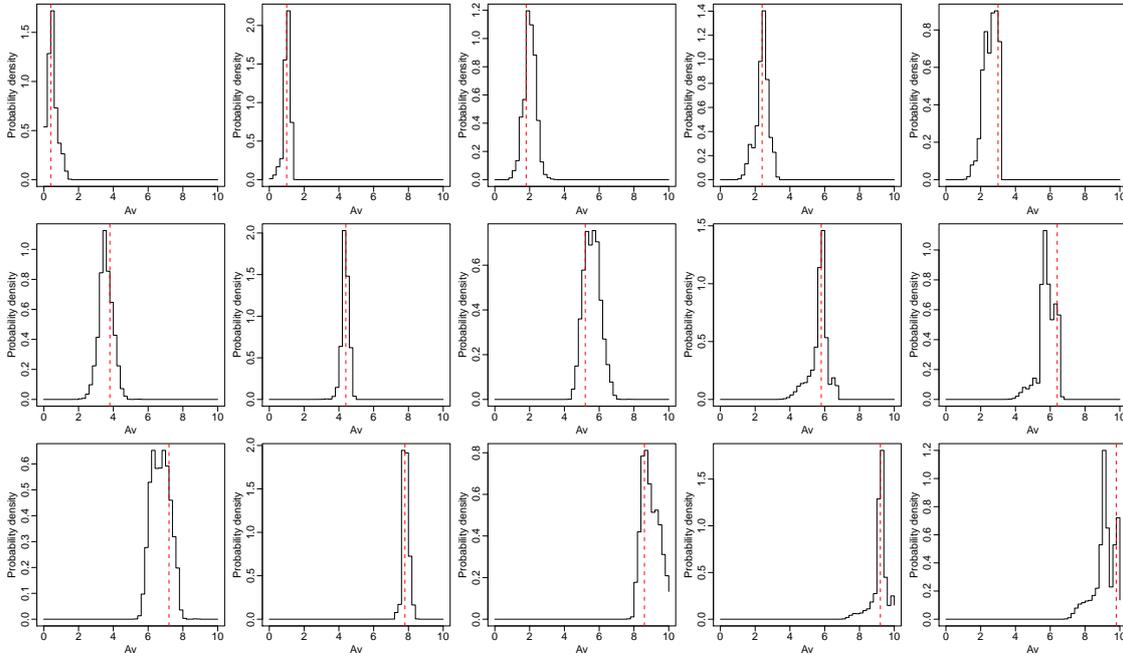


FIGURE 13: Posterior PDF over A_V formed by marginalizing over T_{eff} in Fig. 9.

The effect of an extinction prior on the posterior PDF is easy to predict, of course, and is shown for completeness in Fig. 11 for the 5 kpc case. By construction this slightly lowers the probability of high A solutions. We see relatively little effect, however, because the ridges in the likelihood map do not extend over a very wide range in A_V , i.e. the extinction prior is comparatively weak.

This more or less completes the demonstration. Once we have the posterior PDF, $P(A, T|\mathbf{p}, q)$, we can marginalize over A to get a PDF over T , and vice versa. These are shown for the 5 kpc case with a uniform prior and the age-only HRD prior in Figs. 12 and 13. (They are not very smooth, in part on account of the discrete HRD diagram.) From these we can identify the most probable and secondary solutions, or marginalize to get the expected solution, and we can also calculate corresponding error bars.

5 Extension to other APs and additional data

5.1 Additional data

Imagine we have an estimate for the posterior PDF of the APs ϕ given some data D_1 , $P(\phi|D_1)$. We are now given additional data, D_2 , from which we can also estimate the APs. How do we incorporate these data to estimate the posterior for the combined data, $P(\phi|D_1, D_2)$? We use Bayes' theorem

$$P(\phi|D_1, D_2) = P(\phi|D_1) \frac{P(D_2|\phi, D_1)}{P(D_2|D_1)}. \quad (21)$$

If D_1 and D_2 are independent this becomes

$$P(\phi|D_1, D_2) = P(\phi|D_1) \frac{P(D_2|\phi)}{P(D_2)}. \quad (22)$$

(see also Bailer-Jones & Smith 2010). That is, our previous posterior is updated by multiplying it by the likelihood for the new data $P(D_2|\phi)$ (and dividing by the prior of the new data, which can be found by normalization). In the present context, $P(\phi|D_1)$ is given by equation 13 with $\phi = \{A, T\}$ and $D_1 = \{\mathbf{p}, q\}$. D_2 could refer to the RVS spectrum on Gaia or any other external data, whereby infrared photometry would be particularly relevant for reducing the $T_{\text{eff}}-A_G$ degeneracy (e.g. Knude & Lindstroem 2007).

Note that if D_1 and D_2 are entirely logically dependent (i.e. we re-use the data, even if transformed), then $P(D_2|\phi, D_1) = P(D_2|D_1) = 1$ and so from equation 21 the posterior is unchanged (Jaynes 2003, section 8.12.1). You can't cheat Bayes' theorem.

Knude & Lindstroem (2007) have suggested using the colours $G-BP$, $G-RP$ and/or $BP-RP$ to help determine extinction. If the BP/RP spectrum, \mathbf{p} , has been area normalized then it is independent of G , BP and RP . G is not independent of q (which is $G + 5 \log \varpi$) so we cannot use a colour based on this as D_2 in equation 22. But we could use the colour $BP-RP$, because

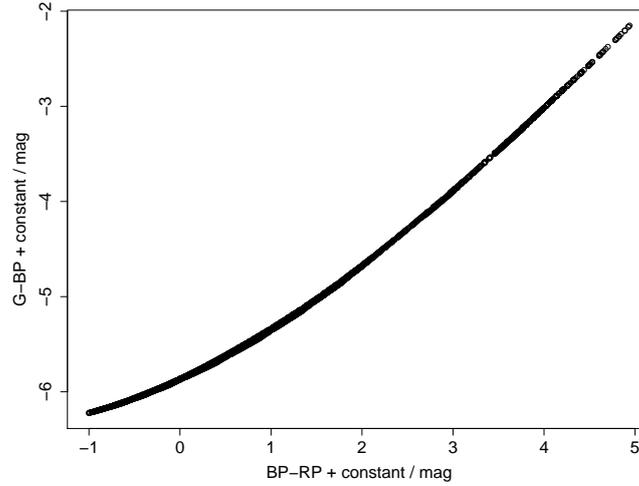


FIGURE 14: Correlation between colours for the noise-free TAG data set at $G=15$, which varies in T_{eff} from 4000-15 000 K, in A_0 from 0–10 mag and $\log g$ from -1.0 dex to +5.5 dex.

it is independent of G . Equivalently, we can incorporate $BP - RP$ as an additional element of \mathbf{p} , because this respects the independence of \mathbf{p} and q assumed in equation 8 and therefore in equation 13. This I did in Bailer-Jones (2010b) and showed that actually $BP - RP$ carries insignificant information beyond BP/RP for determining T_{eff} and A_V .

If we did not use the parallax and HRD then we can of course trivially incorporate $G - BP$ into \mathbf{p} as well and build a new likelihood map. But if we did want to use them and these broad band colours, we would have to derive a new expression for equation 13 in which the spectrum, the colours, G and the parallax all appear as separate terms. However, as the BP and RP magnitudes are almost exactly correlated at a given G , we of course see that $G - BP$ and $BP - RP$ are also highly correlated (Fig. 14), so $G - BP$ does not add any additional information.

5.2 Additional APs

So far I have only considered the APs effective temperature and extinction. Introducing the metallicity results in a dependence of the HRD on this and so the need to modify equation 13. If the metallicity is not degenerate with A and T we could use it to define the HRD for that metallicity, which would have narrower loci. If we can further assume that $[\text{Fe}/\text{H}]$ is independent of A and T when conditioned on the spectrum we can write $P(A, T, [\text{Fe}/\text{H}] | \mathbf{p}) = P(A, T | \mathbf{p}) P([\text{Fe}/\text{H}] | \mathbf{p})$ to permit a simple combination of posterior probabilities. If this conditional independence does not apply, i.e. $[\text{Fe}/\text{H}]$ is degenerate with A and/or T , then we must generalise the likelihood map to include dependence on $[\text{Fe}/\text{H}]$.

The likelihood map (but not the other terms) also assumes that the extinction law (R_0 in equation 18) is constant over the sky, whereas in fact it varies. If we want to try and estimate this

(and initial tests suggest it may be possible from cycle 5 quality GOG data; C. Liu private communication), it is presumably degenerate with A and T also. This would require a rederivation of the theoretical development to accommodate this.

6 Discussion

I have introduced a probabilistic, self-consistent method for incorporating different types of information into the estimation of stellar astrophysical parameters. Being probabilistic it provides not just the single best solution but also a complete posterior probability density function over APs, i.e. we can estimate errors and correlations and identify degeneracies. The approach is self-consistent in the sense that parallaxes and apparent magnitudes are used in a way that they make astrophysical sense, consistent with the types of stars which stellar physics permits. Approaches which consider these data independently could arrive at non-physical solutions, such as very faint early type stars lying at close distances, or very faint M dwarfs at distances which Gaia should not even see. Another advantage of the method is that we can choose which of the information we want to include in the AP estimation without having to change any of the others terms or refit the model. If we do not want to include some piece of information, we just set the distribution over that parameter/data to be flat. Likewise, if we are only interested in one output parameter (e.g. T_{eff}) we can just marginalize over the other ones.

A particular advantage of this approach is that it allows us to make use of parallax in AP estimation without having to go through the difficult problem of generating spectral training data sets from primary stellar AP (mass, age, composition) and then using interpolation of evolutionary tracks (the HRD) to ensure consistency with assigned parallaxes and magnitudes. The only training set which this method relies on is that used to build the likelihood map, which is based only on secondary (atmospheric) APs in which consistency with primary APs is not required. We avoid the problems of inconsistent training data. This is also in line with the ILIUM philosophy of avoiding any interpolation when generating the training sets.

Pattern recognition methods (such as ANNs or SVMs) which try to produce an inverse mapping to the APs of interest from these heterogeneous data are more cumbersome. In general, they do not permit one to simply ignore information we don't have or don't want to use: we instead have to train different models for different input spaces. Most of these methods only give a single solution and are incapable of naturally providing probability distributions (and thus expected uncertainties) over parameters. They likewise cannot or cannot easily incorporate prior information. Most significantly, however, these methods cannot explicitly take into account the constraints of physical background information we have, namely the relationship between T_{eff} , A_V , parallax and magnitude which we know from stellar physics and geometry. This can only be incorporated indirectly via some clumsy tuning of the training data set or transformation of the input variables.

References

- Bailer-Jones C.A.L., 2010a, *The ILIUM forward modelling algorithm for multivariate parameter estimation and its application to derive stellar parameters from Gaia spectrophotometry*, MNRAS, in press
- Bailer-Jones C.A.L., 2010b, *ILIUM V: Further degeneracy mapping, application to the semi-empirical library and the use of the forward model and MCMC to build likelihood maps*, GAIA-C8-TN-MPIA-CBJ-050
- Bailer-Jones C.A.L., Smith K.S., 2010, *Combining probabilities*, GAIA-C8-TN-MPIA-CBJ-053
- Bailer-Jones C.A.L., Smith K.S., Tiede C., Sordo R., Vallenari A., 2008, *Finding rare objects and building pure samples: probabilistic quasar classification from low-resolution Gaia spectra*, MNRAS 391, 1838
- Cardelli J.A., Clayton G.C., Mathis J.S., 1989, *The relationship between infrared, optical, and ultraviolet extinction*, ApJ 345, 245–256
- Cox R.T., 1946, *Probability, frequency, and reasonable expectation*, American Journal of Physics 14, 1–13
- de Bruijne J., 2009, *Gaia astrometric performance: summer-2009 status*, GAIA-CA-TN-ESA-JDB-055-01
- Girardi L., Bressan A., Bertelli G., Chiosi C., 2000, *Evolutionary tracks and isochrones for low- and intermediate-mass stars: From 0.15 to 7 M_{\odot} and from $Z = 0.0004$ to 0.03*, A&AS 141, 371-383
- Gorbikov B., Brosch N., 2010, *Grey Milky Way extinction from SDSS stellar photometry*, MNRAS 401, 231–241
- Jaynes E.T., 2003, *Probability theory. The logic of science*, Cambridge University Press
- Jordi C., 2009, *Photometric relationships between Gaia photometry and existing photometric systems*, GAIA-C5-TN-UB-CJ-041-4
- Knude J., Lindström H., 2007, *Interstellar extinction estimation. Using color-ratios from BP, RP, G, G_{RVS} and JHK photometry*, GAIA-C8-TN-NBI-JK-002-1
- Marshall D.J., Robin A.C., Reylé C., Schultheis M., Picaud S., 2006, *Modelling the Galactic interstellar extinction distribution in three dimensions*, A&A 453, 635–651