ILIUM V:

Further degeneracy mapping, application to the semi-empirical library and the use of the forward model and MCMC to build likelihood maps

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-050
issue:	1
revision:	0
date:	2010-01-24
status:	Issued

Abstract

I present further results of using ILIUM and the forward model for estimating stellar APs from BP/RP spectra (using end-of-mission noise levels). I first show that the T_{eff}–A_V degeneracy is negligible at G=15, unlike at G=18.5. Not surprisingly, including the integrated BP and RP fluxes adds no new information and does not reduce the degeneracy. Second, I show that there is an even stronger [Fe/H]–$\log g$ degeneracy at G=18.5, mostly related to the fact that these APs can hardly be estimated at all at this magnitude. At G=15, in contrast, the "degeneracy" between [Fe/H] and $\log g$ (the likelihood function) can be very well predicted by the analytic expression for the errors in the AP estimation given in CBJ-046. Third, I apply ILIUM trained on synthetic data to the semi-empirical library: systematic results are seen, almost certainly on account of the mismatch between the two data sets, which I discuss. Fourth, I show how the degeneracy map differs from the likelihood map, which corresponds to posterior probability distribution over the APs given a measured spectrum (assuming flat priors on the APs). This gives us the full Bayesian solution to the AP inference (as we can easily adopt arbitrary priors) and is what we should report to give our best estimate of the APs and the corresponding uncertainties. I show how the likelihood map is constructed using the forward model, either by sampling on a regular grid or using a Markov Chain Monte Carlo (MCMC) algorithm.

Document History

Issue	Revision	Date	Author	Comment
1	0	2010-01-24	CBJ	Added likelihood map section (4), "relation" and "context" sections (5 and 7) and modified MCMC section (6).
D	1	2009-12-12	CBJ	First published draft

Contents

TABLE 1: Notation

I	number of bands (pixels in spectrum)
i	counter over band, $i = 1 \ldots I$
p_i	photon counts in band i ($\boldsymbol{p}$ is a spectrum)
C_p	$I \times I$ covariance matrix of $\boldsymbol{p}$
σ_i^2	Diagonal elements of C_p
J	number of APs (astrophysical parameters)
j	counter over AP, $j = 1 \ldots J$
ϕ_j	AP j ($\boldsymbol{\phi}$ is a set of APs)
C_ϕ	$J \times J$ covariance matrix of $\boldsymbol{\phi}$
s_{ij}	sensitivity of band i to AP j, $\frac{\partial p_i}{\partial \phi_j}$
S	sensitivity matrix, $I \times J$ matrix with elements s_{ij}
$f_i(\boldsymbol{\phi})$	forward model for band i
A_V	$= A_0$, the extinction parameter
$\log$	base 10 logarithm
$\ln$	natural logarithm

1 Introduction

In Bailer-Jones (2009a,b,c,d) I introduced the ILIUM algorithm and showed how it can be used to estimate stellar APs from Gaia BP/RP spectra. (For an overview see the MNRAS paper – Bailer-Jones (2010a), hereafter CBJ10a.) Here I present some additional results from using this algorithm and from using the forward model directly. Several results concern the "degeneracy map", which is a way of showing how a measured spectrum (and corresponding AP solution found by ILIUM) is consistent given the noise with forward model-predicted spectra corresponding to other APs. The map shows probability contours as function over two APs, holding other APs constant (see CBJ10a section 6.5 for an explanation.) This is not a delta function because of the errors in the measurements, and these give rise to a degeneracy between APs.

Here I show degeneracy maps for both T_{eff}–A_V (section 2.1) and [Fe/H]–$\log g$ (section 2.3) at a range of magnitudes. In the former section I also show the unsurprising result that including the integrated BP and RP fluxes in the likelihood function adds no information beyond the resolved BP/RP spectrum, and therefore does not help to reduce the T_{eff}–A_V degeneracy. In section 3 I apply ILIUM fitted on synthetic data to the semi-empirical library in order to examine the data mismatch problem, and thus take a small step toward the "blind" application situation we will have with real Gaia data.

The degeneracy map is just one way of showing possible confusion between solutions. In section 4 I introduce the likelihood map, which shows the posterior probability distribution over the APs assuming uniform (flat) priors. It is different in principle and in practice from the degeneracy map and in many ways provides a more appropriate summary of probable solutions. In section 5 I demonstrate that for small errors the analytic approximation to the uncertainties

in the ILIUM AP estimates agree with empirical calculations of the degeneracy map, i.e. the degeneracies can be predicted. Finally, in section 6, I demonstrate how the Markov Chain Monte Carlo (MCMC) can be used to sample the likelihood map.

In the interests of brevity this TN assumes familiarity with CBJ10a and the notation therein (summarized in Table 1). The datasets/problems used here are a subset of those introduced in that paper:

- TAG: Estimation of $(T_{eff}, A_V)+\log g$ (2 strong and 1 weak AP), for stars with [Fe/H] = 0. This has 2740 stars, of which a random selection of 1000 is used for evaluation.

- TGM: Estimation of $T_{eff}+(\log g, [Fe/H])$ (1 strong and 2 weak APs), for stars with $A_V = 0$. This has 4361 stars of which a random selection of 1000 is used for evaluation.

To be consistent with the other TNs in this series, I continue to use the symbol A_V to indicate the extinction parameter, not the extinction in the V band. CU8 now uses the symbol A_0 for the extinction parameter in order to avoid this ambiguity.

2 Further examples of degeneracy and degeneracy maps

2.1 Degeneracy mapping with the TAG problem

This problem was fully described in CBJ10a (section 6.5), where a degeneracy map for T_{eff} and A_V at fixed $\log g$ and [Fe/H] at G=18.5 was presented. This is reproduced in Fig. 1. Informally speaking, given a star with APs show by the red cross, any other star (AP combination) lying within the contours has an expected spectrum which is indistinguishable from this to within the expected noise. In other words, there is an intrinsic degeneracy in the spectra. Clearly the extent of the degeneracy depends on the noise. Figs. 2 and 3 show the same at G=20 and G=15 respectively. At G=20 the degenerate region is naturally larger, and at G=15 is smaller. Indeed, it is so small that it is no longer sampled sufficiently by the density of the grid used to build the degeneracy map (0.2 mag in A_V and 0.01 dex in $\log(T_{eff})$), so it is not useful to call this a degeneracy. Somewhere between G=15 and G=18.5 a significant degeneracy sets in.

It has been suggested that use of the BP−RP colour could help to break this T_{eff}−A_V degeneracy. Knude & Lindstroem (2007) show that *if* the spectral type (T_{eff}) is known, then a combination of the G band flux and the BP−RP colour can determine A_V. But because of the degeneracy T_{eff} will *not* be known independently of A_V (if the inference uses only the BP/RP spectrum). As the BP−RP colour is derived by integrating this spectrum, it does not add any additional information, so it's difficult to see how it could reduce the degeneracy. The only possibility is if ILIUM is not making efficient use of the BP/RP spectrum. We can test whether BP and RP fluxes add information using the degeneracy map, because it only uses the forward model and not the ILIUM search algorithm.
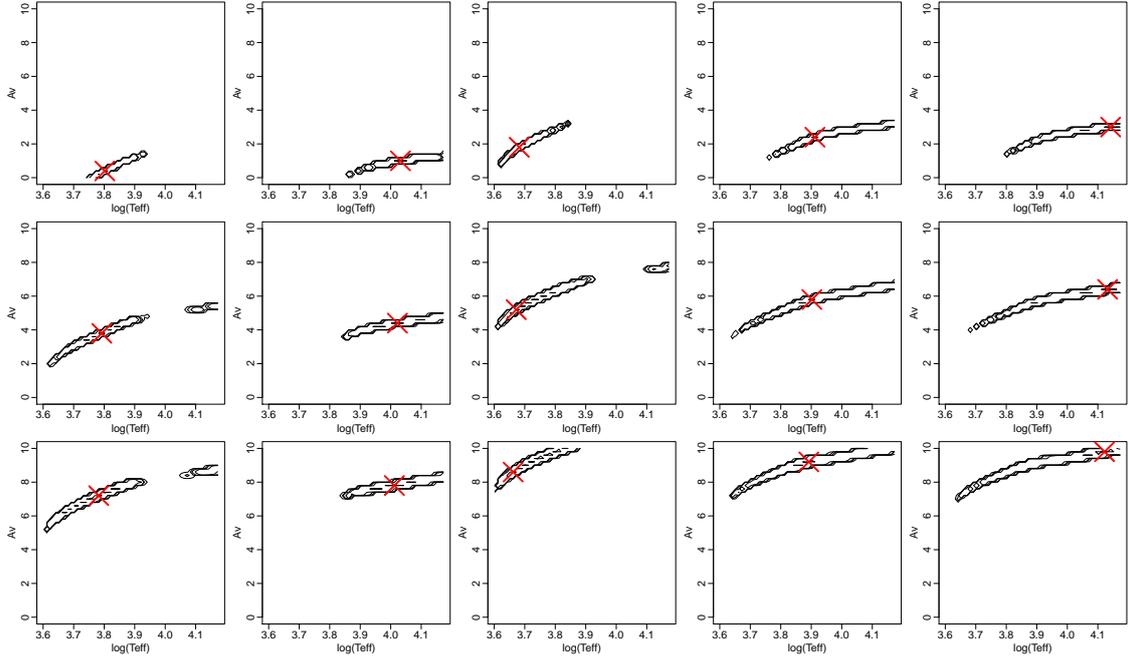
FIGURE 1: Degeneracy map for A_V and T_{eff} on the TAG problem for stars with $\log g = 4.0$ dex (and solar metallicity) at G=18.5. Contours are at $\log P_{lim} = \{-4, -3, -2, -1\}$. From Bailer-Jones (2010a)
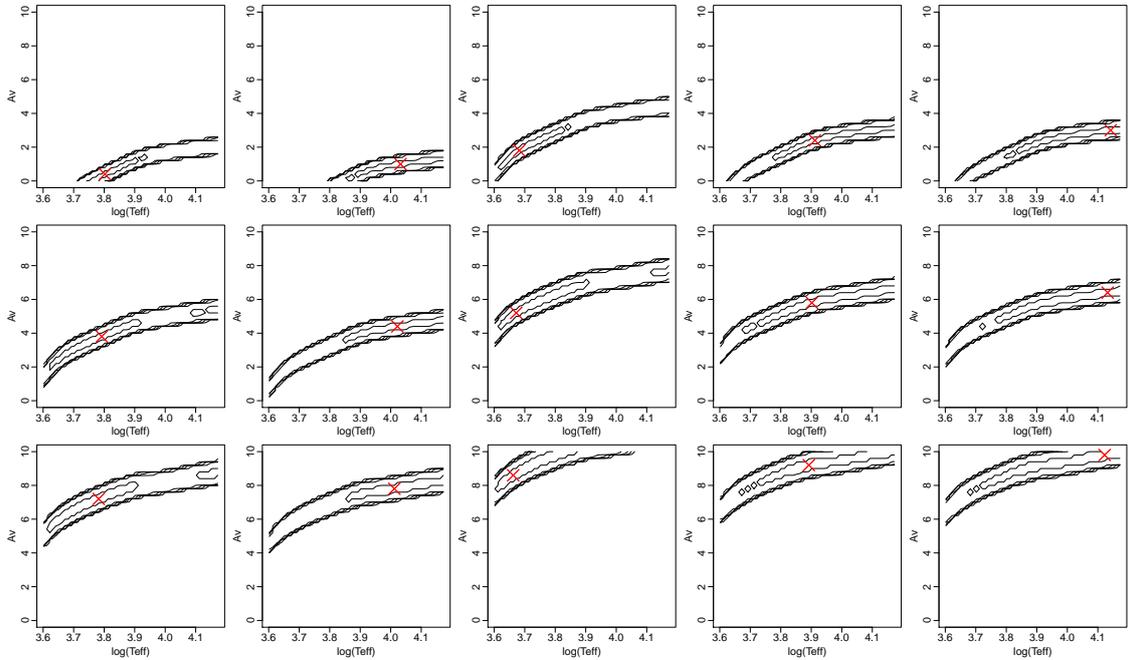
FIGURE 2: As Fig. 1 but for G=20
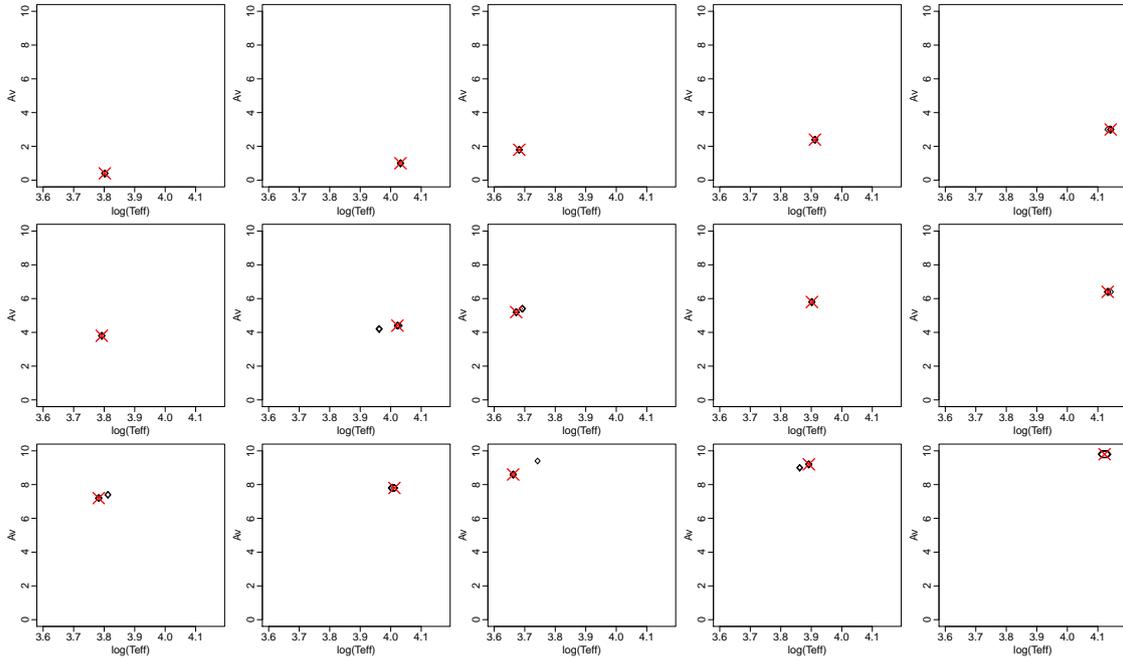
FIGURE 3: As Fig. 1 but for G=15

Fig. 4 shows the degeneracy map constructed using just two bands, the integrated BP and RP fluxes, for exactly the same problem show in Fig. 1. As this is at constant G there is effectively only one independent measurement here, so it is not surprising that alone these fluxes hardly constrain T_{eff} or A_V (although at low extinction it is not much worse than the G=20 result using the full spectrum). I now add these two bands to the full spectrum and recalculate the map, shown in Fig. 5. Comparing this with the map based only on the spectrum (Fig. 1), we see virtually no difference. As expected, the integrated BP and RP fluxes add no information beyond the resolved spectrum and do not help to reduce the degeneracy.

2.2 Application to the TGM problem

This problem was only briefly reported on in CBJ10a (section 7 and Table 2). It is the problem in which we have a 1D strong component over T_{eff} and a 2D weak component over $\log g$ and [Fe/H] (and I fit it on zero extinction data). It is therefore relevant to situations where we can be confident that extinction is zero (or incur little net penalty by making this assumption), e.g. at high Galactic latitudes. While I reported results in CBJ10a, no plots were shown. Figs. 6 and 7 show cuts through the forward model.
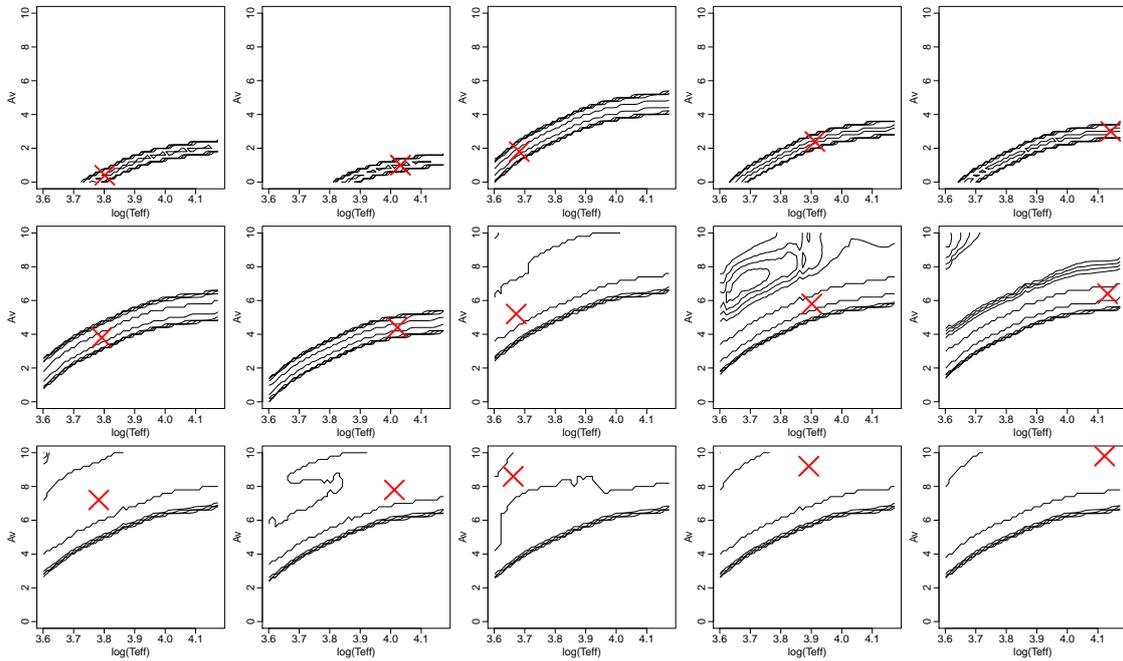
FIGURE 4: As Fig. 1 but using only the integrated BP and RP flux to determine D^2.
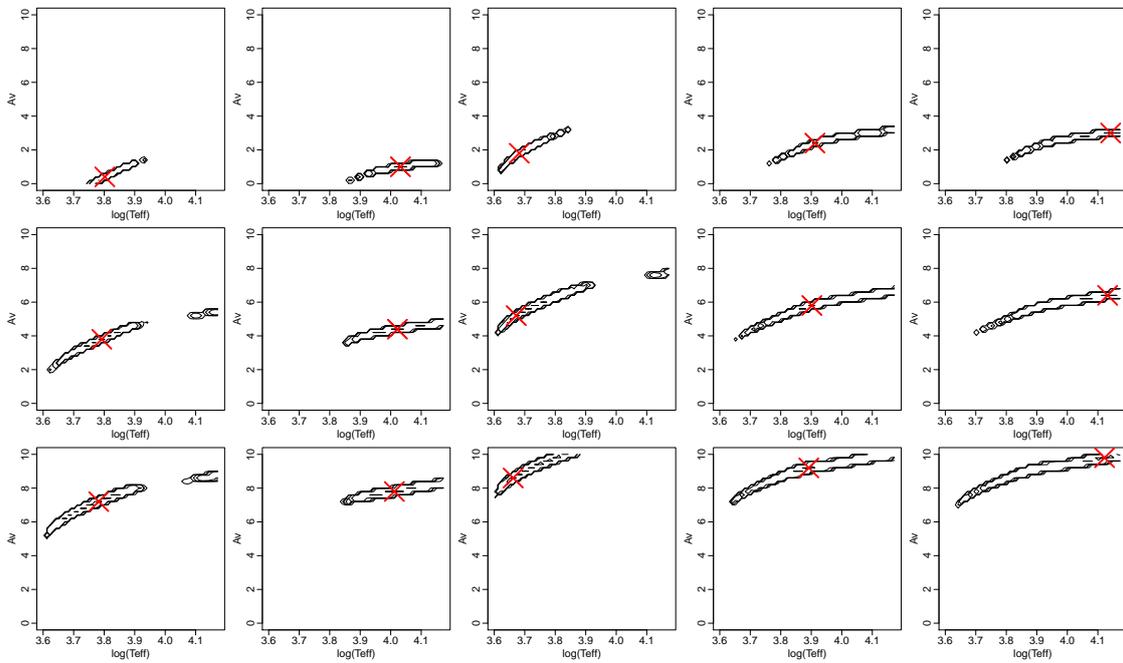
FIGURE 5: As Fig. 1 but using both the BP/RP spectrum (as in that plot) and the integrated BP and RP fluxes to determine D^2.
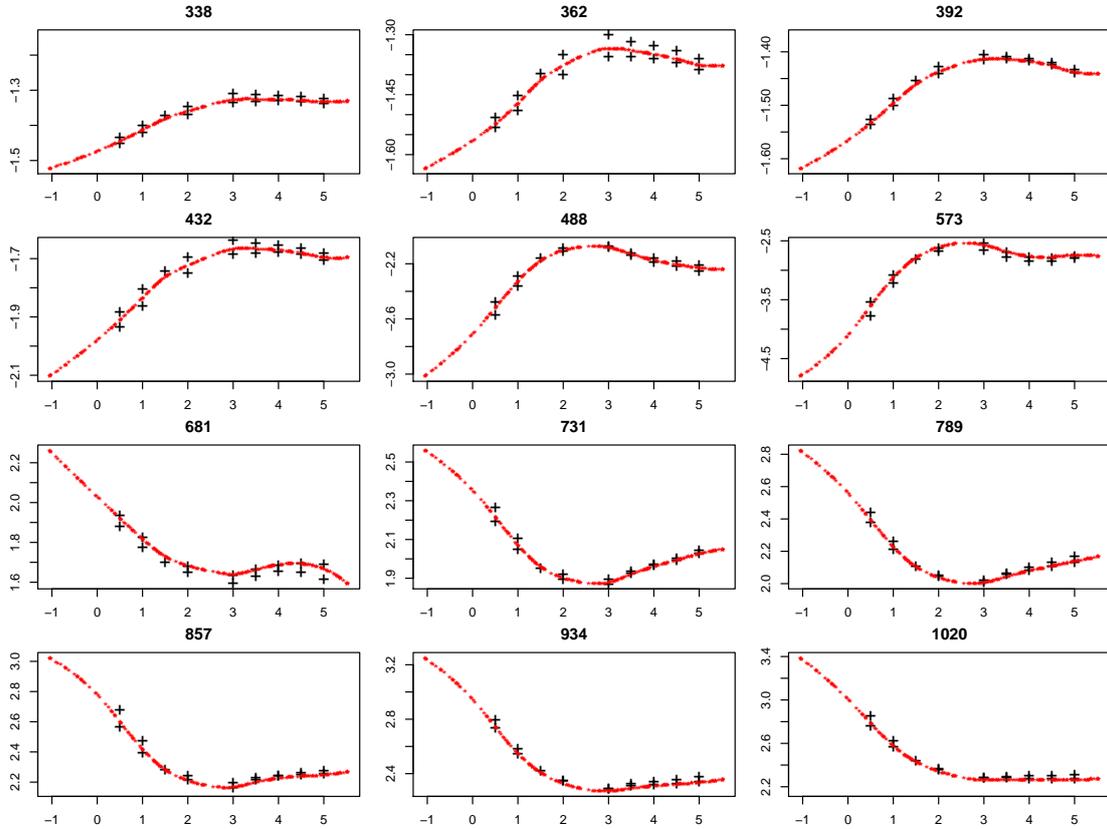
FIGURE 6: 1D slice through the full 3D forward model on the TGM problem for 12 different bands (the central wavelength of which in indicated at the top of each panel in nm). Each panel shows the variation in flux as a function of $\log g$ for [Fe/H] fixed at -2 dex and T_{eff} fixed at 4000 K. The black crosses are the (noise-free) grid points, the small red stars are the forward model predictions (at randomly selected AP values). Double points at some $\log g$ points correspond to different [α/Fe] values.

2.3 Degeneracy mapping with the TGM problem

Fig. 8 shows a degeneracy map between $\log g$ and [Fe/H] for stars at 6000 K at G=18.5. It was show in CBJ10a that estimates of these two APs are rather poor at G=18.5. It should come as no surprise, therefore, that there is a very large degeneracy between these APs across the whole grid. At G=15 (Fig. 9) the situation is much better. The errors are correlated, but the regions are compact and reasonably symmetric. We will see in section 5 that this degeneracy pattern is approximated well by the analytic approximation for the ILIUM uncertainties.
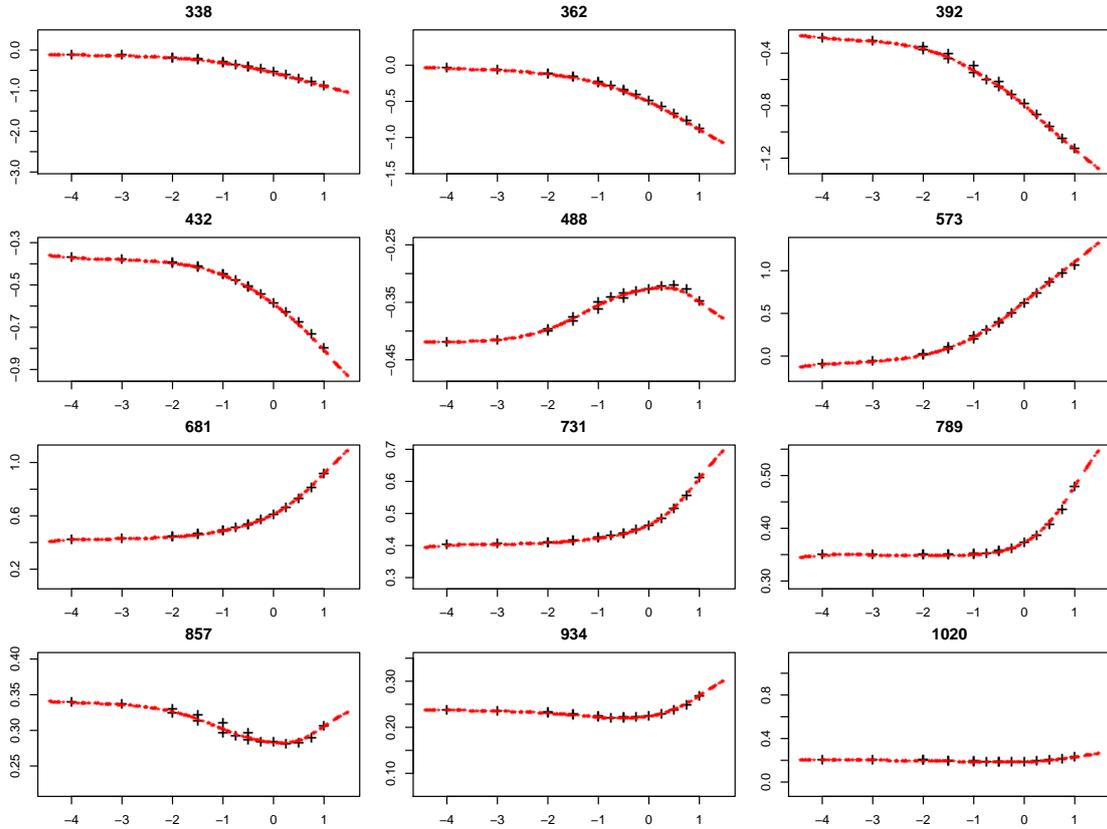
FIGURE 7: As Fig. 6 but for the fluxes as a function of [Fe/H] for $\log g$ fixed to 3.0 dex and T_{eff} fixed at 6000 K.

3 Application of the TGM ILIUM model to the semi-empirical library

Until now, most AP estimation algorithms in CU8 have been built using synthetic spectra, and their performance assessed on similar synthetic spectra drawn from the same grid. But real spectra are not synthetic spectra. There may be significant differences both in how the true fluxes vary with the true APs and in the Gaia instrument model compared to the real instrument, and there will be unmodelled cosmic variance in real spectra which is absent in the synthetic grids. These differences I collectively refer to as the *data mismatch problem.* This adversely affects both the classification reliability and the accuracy of our performance predictions. We could try to mitigate the former by training the models on spectra drawn from the real Gaia sample, although the substantial issue of having to assign accurate APs to these training data remains. (One approach, namely using a small set of AP reference stars to calibrate the synthetic grids, is explained in Bailer-Jones 2009e.)

For this reason CU8 has constructed a semi-empirical library of stellar spectra (Tsalmantza & Bailer-Jones 2009). To build this the authors took a set of high quality SDSS spectra (with
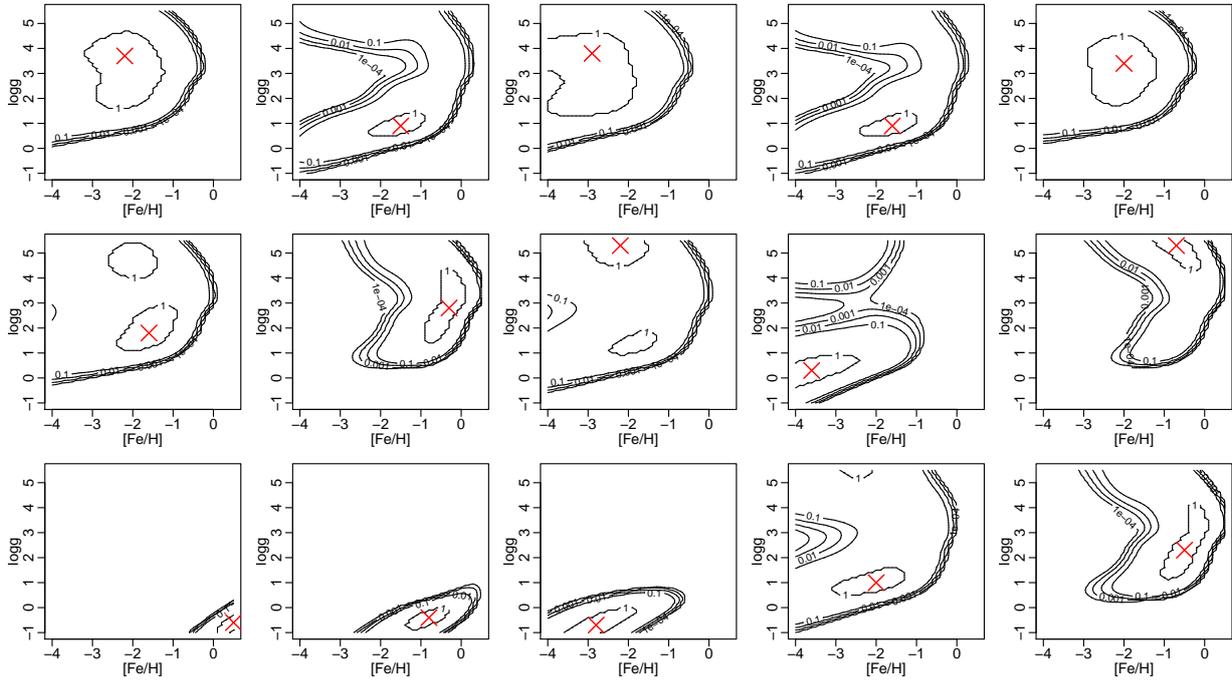
FIGURE 8: Degeneracy map for [Fe/H] and $\log g$ on the TGM problem for stars with $T_{\text{eff}} = 6000\,\text{K}$ (and zero extinction) at G=18.5.
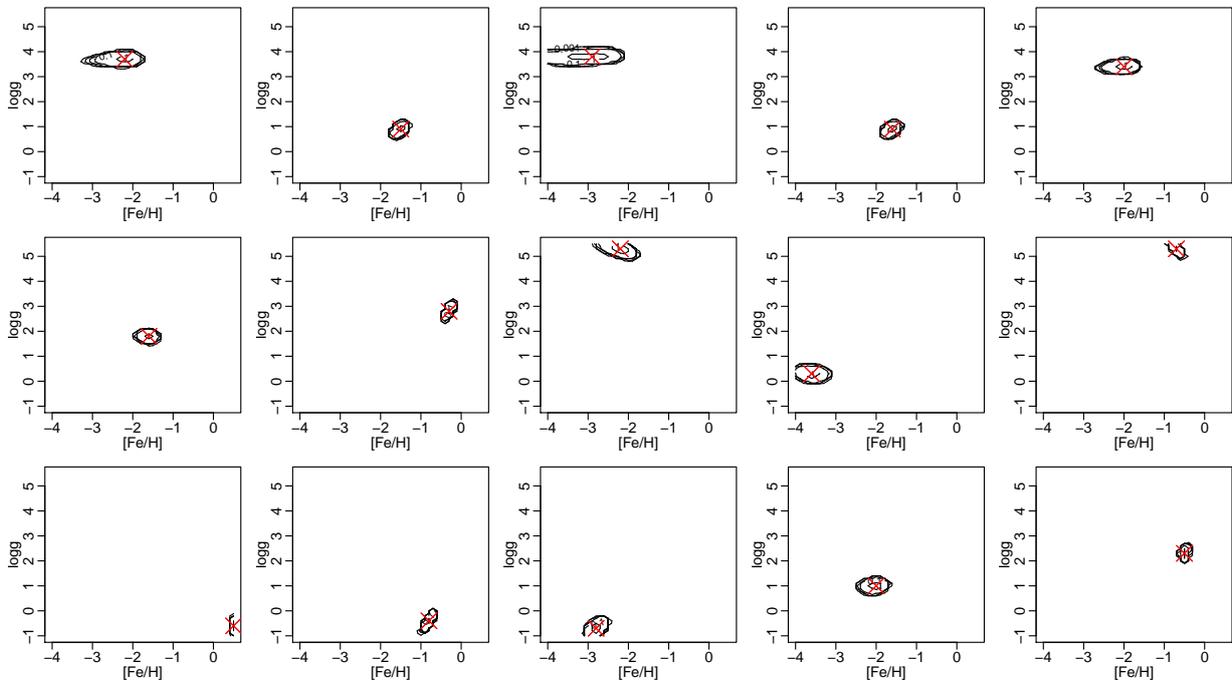
FIGURE 9: As Fig. 8 but for G=15.
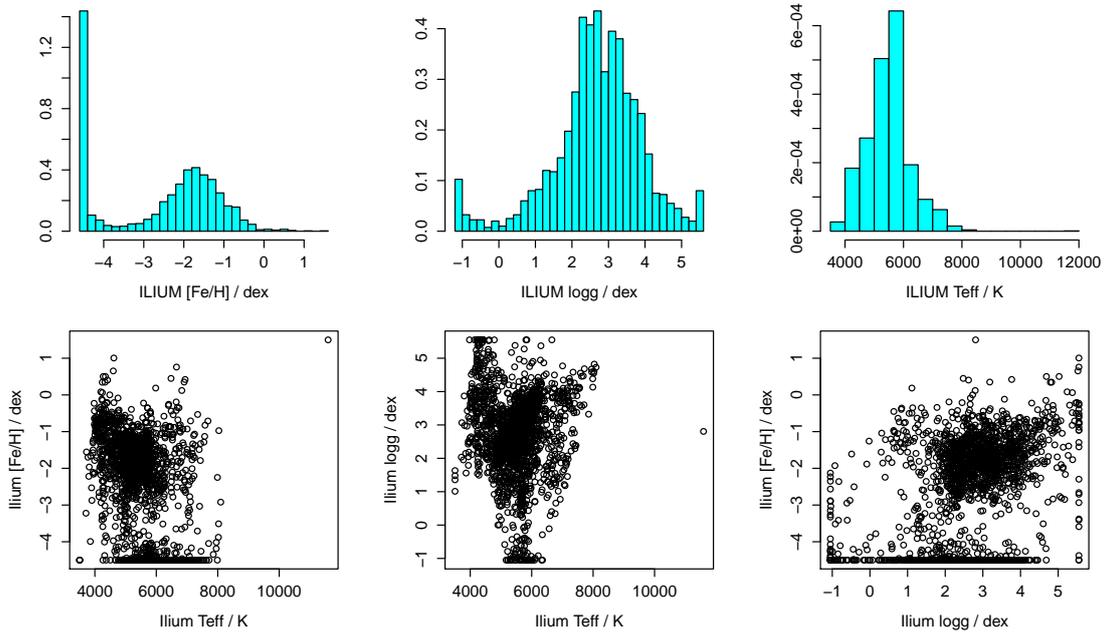
FIGURE 10: Distribution of estimated APs in the semi-empirical library problem

APs available from the SDSS SEGUE Stellar Parameter Pipeline, SSPP). They fit the common wavelength region of each spectrum to the synthetic library to find the best match (via a masked χ^2 minimization), and then used this match both to extend the wavelength range to the full BP/RP range and to assign APs to the SDSS spectra. BP/RP spectra of these have then been simulated by GOG.

Here I use ILIUM fit on synthetic data to estimate the APs of the semi-empirical library. This emulates the kind of data mismatch problem we will face when we come to estimating APs of real Gaia data. As the SDSS stars were selected to have low extinction ($E(B - V) < 0.3$), I use ILIUM fitted on the TGM grid to estimate $T_{\rm eff}$, $\log g$ and [Fe/H] assuming the extinction to be zero. Note that these training data are GOG cycle 3 simulations whereas the semi-empirical library data is modelled using GOG cycle 6. Between these two cycles there were changes in both the dispersion function and the flux scale of BP/RP, as described in Mantis issue 5656. The plots attached to that issue show that the flux scale differs systematically by about 20% and the wavelength centres of the pixel by up to 0.1 pix. The former is huge, but I remove it in this application of ILIUM by performing an area normalization on all spectra (train and validation sets). The wavelength mismatch is relatively small (although larger than any wavelength calibration error we should encounter with real Gaia data), so I simply neglect it.

The results of this model, applied to G=15 data, are shown in Figs. 10, 11 and 12. Recall that ILIUM is permitted to extrapolate modestly beyond its training grid, so can assign APs over the following ranges: [Fe/H] $= [-4.5, +1.5]$ dex, $\log g = [-1.05, +5.55]$ dex, $T_{\rm eff} = [2900, 16100]$ K. Looking first at the $T_{\rm eff}$ estimation in Fig. 10, we see that ILIUM only assigns low $T_{\rm eff}$ and
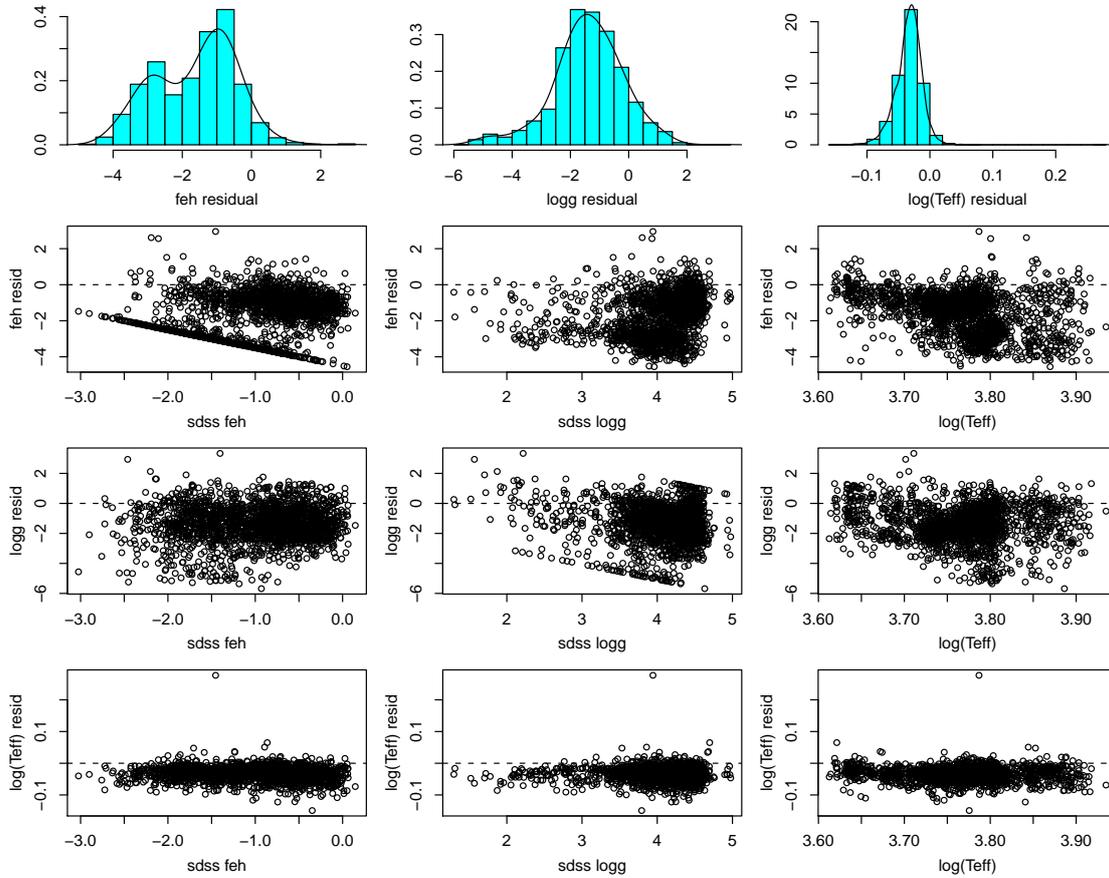
FIGURE 11: AP residuals – ILIUM minus SDSS estimates – for the semi-empirical library problem

does not make any assignments at its limits. This agrees with the selection limits imposed by SSPP, which only tries to assign APs to cool stars. The SSPP APs are assigned using a range of methods and spectral libraries independently of ILIUM. Although not a major test, this is a positive confirmation that ILIUM achieves a plausible T_{eff} distribution for this strong AP.

The situation is somewhat different for $\log g$ and especially for [Fe/H]. Here we see a large number of stars at the very lowest [Fe/H] limit (and to a lesser extent at both $\log g$ limits). Given our secure knowledge of the rarity of very metal poor stars (this is a prior!), this result is not plausible and suggests that there is a signature in the semi-empirical data which ILIUM is identifying as a signal of low metallicity. This may be the consequence of unmodelled cosmic variance or some other data mismatch.

Figs. 11 and 12 show the residuals plotted as a function of the SDSS APs and library APs respectively. These plots should not be considered as showing "errors", because neither SSPP nor the library provides unequivocal truth: they are simply *other* estimates of the APs of SDSS spectra (although the former are based on much higher resolution data, multiple semi-independent
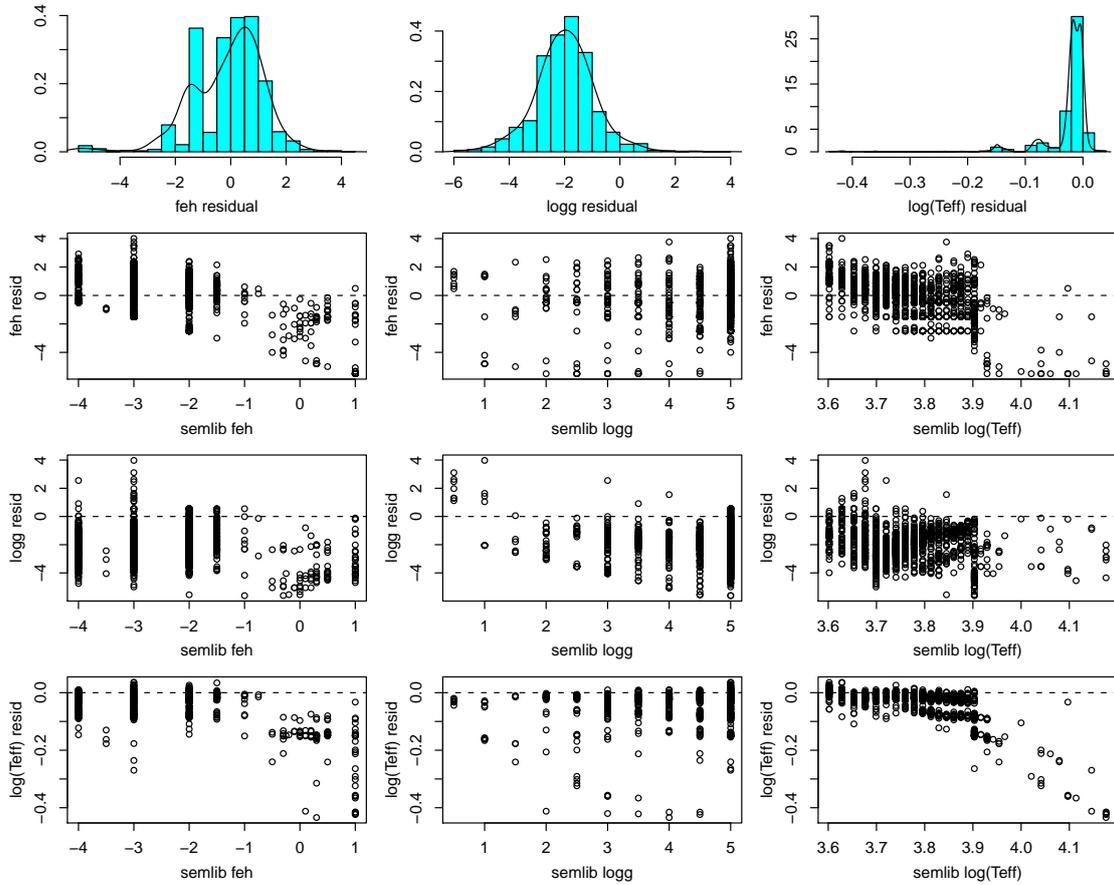
FIGURE 12: AP residuals – ILIUM minus semi-empirical library estimates – for the semi-empirical library problem

methods, and a concerted effort over several years). We of course again see the bogus extreme assignments of $\log g$ and [Fe/H] by ILIUM, here as the strong diagonals in the relevant panels of Fig. 11. But we in addition see a more subtle and global systematic difference, in the sense that the residuals are systematically offset from zero. This does not necessarily indicate an "error" in any of the AP assignments. It is more likely to reflect a different AP calibration between the Gaia synthetic spectra and the SSPP AP system. We also see this systematic in Fig. 12, which cannot be for exactly the same reason because both the semi-empirical library and ILIUM APs are based on the same synthetic library. On the other hand, the *assignment* of the APs in the semi-empirical library *is* affected by the mismatch problem, because it affects which spectrum is chosen as the closest match in assigning APs, so it comes down to the same cause. As the systematic is a simple offset independent of the APs, it could be corrected for. (If the systematic error were a function of the APs, then generally it could not be corrected for, as explained in appendix A of CBJ-043.)

Future work should examine more closely the reasons for the systematic differences as well as the bogus extreme AP assignments. It is likely that the assumption of exactly zero extinction

for the SDSS spectra is wrong. This is a strong AP, so even a small amount of unmodelled extinction could be interpreted by ILIUM as the signal of another AP, e.g. of low metallicity. Perhaps this accounts for the some AP estimations piling up at the extremes. We should also repeat the work with data from the same GOG model, just to ensure there are no significant instrument model mismatch issues.

4 The likelihood map (AP probability distribution) and how it differs from the degeneracy map

The degeneracy maps discussed above show which stars have a spectrum consistent (given the noise) with the expected spectrum according to the χ^2 metric. That is, given the expected spectrum $\boldsymbol{p}(\boldsymbol{\phi})$ of a star with APs $\boldsymbol{\phi}$, we identify all those stars, $\boldsymbol{p}^\star(\boldsymbol{\phi}^\star)$, which have a spectrum such that the squared distance

$$
\begin{aligned}
D^2 &= \delta\boldsymbol{p}^T C_p^{-1} \delta\boldsymbol{p} \\
&= \sum_{i=1}^{i=I} \left(\frac{p_i - p_i^\star}{\sigma_i} \right)^2
\end{aligned}
$$

is less than some threshold. In this equation $\delta\boldsymbol{p} = \boldsymbol{p} - \boldsymbol{p}^\star$ and C_p is the covariance matrix (expected noise in $\boldsymbol{p}$) which is diagonal with elements σ_i^2. D^2 has a χ^2 probability density function (PDF), $P'(D^2 \,|\, \boldsymbol{\phi})$, with I degrees-of-freedom, so we can easily set the threshold so that it equals a specified probability. As $P'(D^2 \,|\, \boldsymbol{\phi})$ is a *density* function,[1] the probability of measuring any exact distance is infinitesimally small. So in the usual way of orthodox statistics, we integrate up $P'(D^2 \,|\, \boldsymbol{\phi})$ from zero until we get to some threshold value, D_{lim}^2 such that the integrated probability is P_{lim} (see left panel of Fig. 13). Thus all stars $\boldsymbol{p}^\star(\boldsymbol{\phi}^\star)$ which have a squared distance $D^2 = D_{\text{lim}}^2$ have a probability P_{lim} of having a spectrum indistinguishable from $\boldsymbol{p}(\boldsymbol{\phi})$ (under the null hypothesis of Gaussian noise). To produce the degeneracy map we in principle search the infinite continuous set of all stars and find those which have a distance corresponding to P_{lim}. In practise we search on a discrete grid and join them with a contour. The degeneracy maps in this TN show contours with $\log P_{\text{lim}} = \{-4, -3, -2, -1\}$.

Put simply, for a star with *given* APs $\boldsymbol{\phi}$ and expected spectrum $\boldsymbol{p}(\boldsymbol{\phi})$, the degeneracy map shows (in AP space) all stars which are degenerate within the specified P_{lim}. For example, the stars within the contour $P_{\text{lim}} = 0.01$ are all those stars which *individually* have a probability of more than 1% of showing a spectrum like $\boldsymbol{p}(\boldsymbol{\phi})$. This does *not* mean that these are 99% of all stars which could have a spectrum like $\boldsymbol{p}(\boldsymbol{\phi})$.

This difference is very important. While the degeneracy map shows which stars have the same spectrum (to within some noise level) as $\boldsymbol{p}(\boldsymbol{\phi})$, it does not tell us about the relative probabilities of all these possible solutions. It cannot, because this would require us to normalize over the

---

[1]In this section I use the ′ symbol to indicate a probability *density* function, P', in order to distinguish it from an integral of this, P, sometimes called the probability *distribution* function.
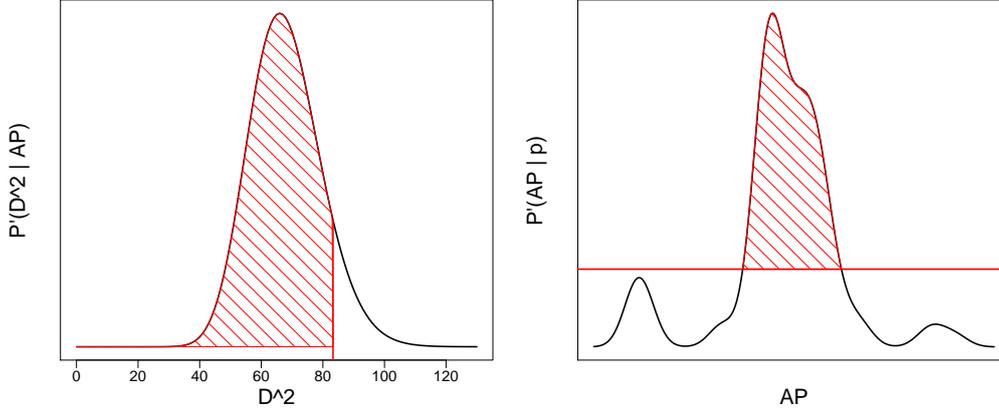
FIGURE 13: Illustration of the difference between how we calculate the degeneracy map (left) and how we calculate the likelihood map (right). For the former we find all those stars at a squared distance D^2 which have $\int_0^{D^2_{\text{lim}}} P'(D^2 \,|\, \phi) \, dD^2 = P_{\text{lim}}$ (here shown for 0.9) and then plot them in the AP space as a contour. To calculate the likelihood map we find all those stars which have $\int_V P'(\phi \,|\, p) \, d\phi = P_{\text{lim}}$ (shown here schematically), where V is the volume of AP space which contains the fraction P_{lim} of the largest probability densities.

AP space, whereas the calculation of D^2 does not involve any normalization which takes into account the other degenerate solutions.

When inferring APs from a spectrum, what we really want to know is $P'(\phi \,|\, p)$, the posterior probability density function over the APs *given* the (noisy) observed spectrum, p. The peak of this is the single most probable solution, the mean is the expected solution. More generally we want to find the confidence intervals, or rather the set of (for example) 99% most probable solutions. To do this we make a cut through the density function such that 99% of the probability mass of the function lines above the line (see right panel of Fig. 13). We can do this if $P'(\phi \,|\, p)$ is properly normalized, that is *normalized over all the APs*: $\int P'(\phi \,|\, p) d\phi = 1$. This is completely different from the degeneracy map, which uses $P'(D^2 \leq D^2_{\text{lim}} \,|\, p^\star(\phi))$ normalized over all D^2 *for each star individually*, i.e. $\int P'(D^2 \,|\, p^\star(\phi)) \, dD^2 = 1$. These two concepts are different in the same way that $P(\phi|p)$ is not the same as $P(p|\phi)$. This is a crucial difference of practical importance, and not just an academic distinction, as we will now see.

To calculate the posterior probability density $P'(\phi \,|\, p)$ we use Bayes' theorem

$$P'(\phi \,|\, p) \;=\; \frac{P'(p \,|\, \phi) P'(\phi)}{P'(p)} \; . \tag{3}$$

The denominator can be interpreted as a normalization constant for a given measurement, i.e. the integral of the numerator over all APs, $P'(p) = \int P'(p \,|\, \phi) P'(\phi) d\phi$. The term $P'(p \,|\, \phi)$ is the *likelihood function*, and gives the probability of observing the spectrum given the APs. If there were no noise, then this probability would be a delta function at the one and only possible spectrum. But the measured spectrum is noisy, $p = p^0(\phi) + \epsilon$, where p^0 indicates the noise-free

expected spectrum and $\boldsymbol{\epsilon}$ is the noise, so the probability function has a finite width. Making the same assumptions as were used for the degeneracy map, we model each element of $\boldsymbol{\epsilon}$ with a zero mean Gaussian distribution, so that

$$P'(\boldsymbol{p} \,|\, \boldsymbol{\phi}) \;=\; \frac{1}{\sqrt{2\pi|C_p|}} \exp\left(-\frac{1}{2}\delta\boldsymbol{p}^T C_p^{-1} \delta\boldsymbol{p}\right) \qquad (4)$$

where $\delta\boldsymbol{p} = \boldsymbol{p} - \boldsymbol{p}^0(\boldsymbol{\phi})$ and C_p is the covariance matrix of $\boldsymbol{p}$ (and hence of $\boldsymbol{\epsilon}$). As before, if the noise in each pixel is independent then $C_p = \mathrm{diag}(\{\sigma_i^2\})$ and this reduces to

$$P'(\boldsymbol{p} \,|\, \boldsymbol{\phi}) \;=\; \frac{1}{\sqrt{2\pi|C_p|}} \exp\left(-\frac{1}{2}D^2\right) \qquad (5)$$

where

$$D^2 = \sum_{i=1}^{i=I} \left(\frac{p_i - p_i^0}{\sigma_i}\right)^2 \;. \qquad (6)$$

We can now use equation 5 to calculate the probability that the spectrum we observe, $\boldsymbol{p}$, could have been produced under the noise model by a star with noise-free spectrum $\boldsymbol{p}^0(\boldsymbol{\phi})$ with APs $\boldsymbol{\phi}$. Substituting this into equation 3 and adopting a suitable prior, $P'(\boldsymbol{\phi})$, over the APs, we can calculate the posterior probability distribution for $\boldsymbol{\phi}$. We could do this most simply by calculating $\boldsymbol{p}^0(\boldsymbol{\phi})$ on a dense grid using the ILIUM forward model and plotting as a function of $\boldsymbol{\phi}$ as we did for the degeneracy map. If we do this using a flat, uninformative prior, this gives what I shall call the *likelihood map*. As the prior is flat, the likelihood map is identical to the likelihood function to within a multiplicative factor.

Note that the ILIUM AP estimation algorithm is not used to calculate the likelihood map. We just use the ILIUMforward model to predict the spectrum on a regular grid of APs and use the above equations to assess their probability.

Fig. 1 shows the degeneracy map for the TAG problem at G=18.5. Fig. 14 shows the corresponding likelihood map for the same data. The most obvious difference is that the region of AP space occupied by the 1% contour in the likelihood map is much smaller than the 1% contour in the degeneracy map. The reason is that for a *given* expected spectrum, there is a relatively large region of AP space which yields degenerate spectra (to a 1% confidence level), whereas for a *given* measured spectrum the region of space occupied by just the "closest" 99% of AP solutions is much smaller. This is a conceptual difference arising from what we normalize with respect to. Note that this difference has nothing to do with priors (the prior is flat in the likelihood map).

A second difference is that the degeneracy map looks smoother than the likelihood map. A priori there is no reason to expect that the posterior PDF in the likelihood map should decrease monotonically with some Euclidean distance in AP space. (The same can be said for the degeneracy map.) The two maps show quite different things and it is not obvious whether or why one should be smoother than the other. It might be that in the likelihood map we are probing much smaller volumes of the possible AP space, so we are more sensitive to the discrete grid used for the plotting.
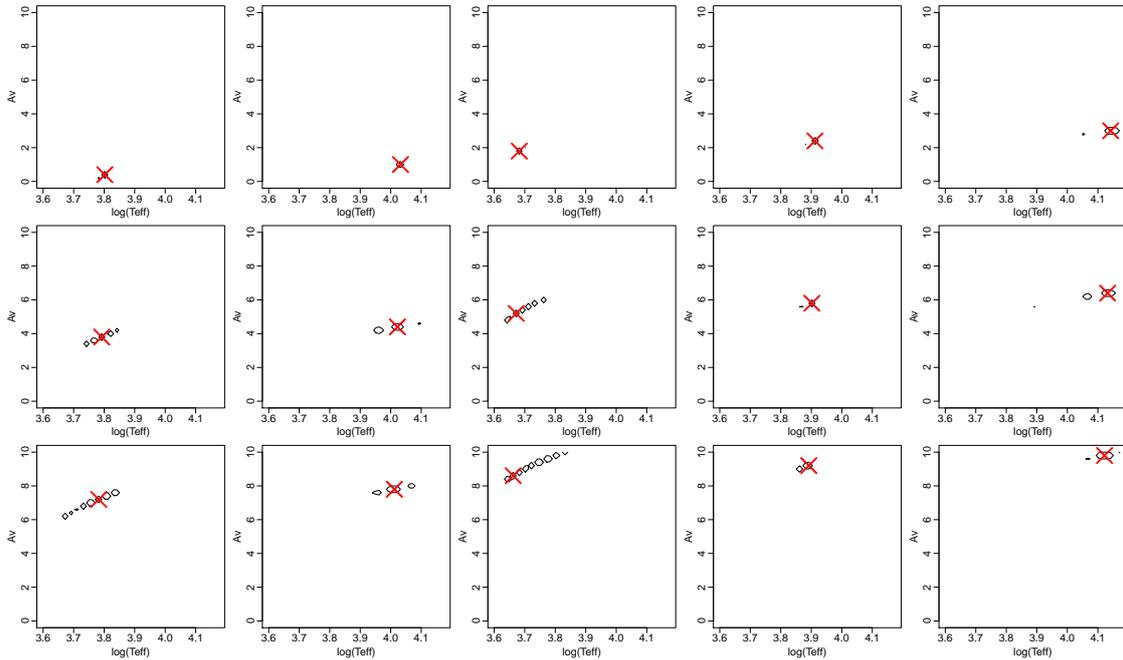
FIGURE 14: Likelihood map corresponding to the degeneracy map show in Fig. 1 (G=18.5). The contour includes the 99% most probable solutions in the way shown in the right panel of Fig. 13
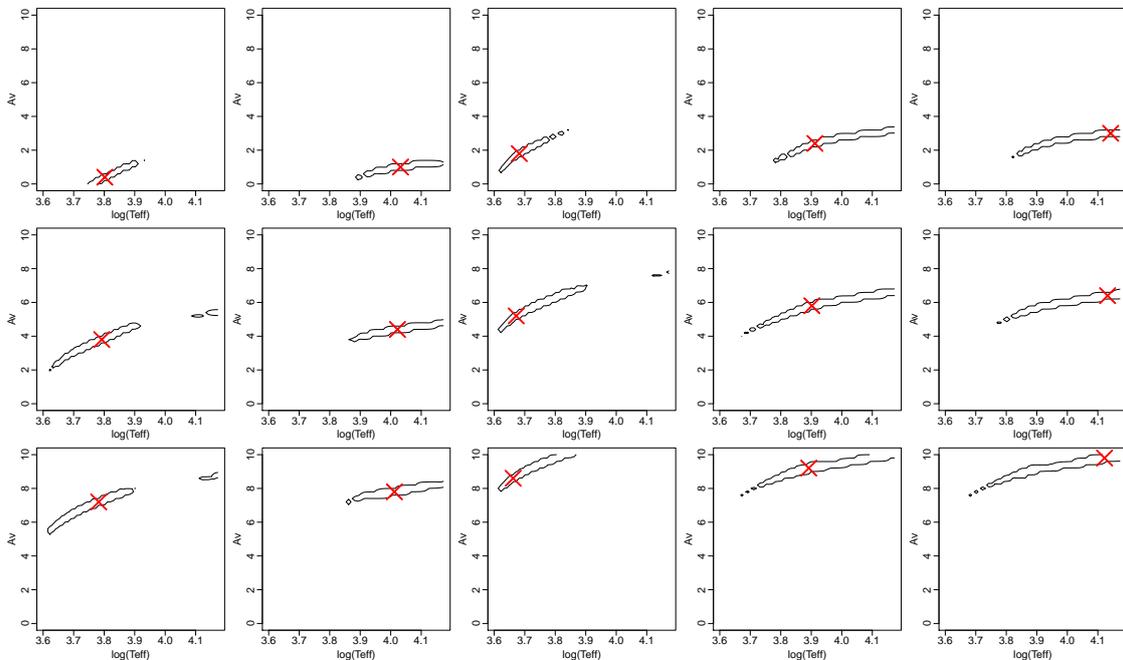
FIGURE 15: As Fig. 14 but for G=20. Compare with the corresponding degeneracy map in Fig. 2.

Fig. 15 shows the likelihood map for the same stars but now for noise corresponding to observations at G=20. Naturally the region containing 99% of the AP solutions is larger than at G=18.5. If we compare this with the corresponding degeneracy map (Fig. 2) we again see that the likelihood map is more compact, although the contrast is less strong than at G=18.5. (Marginalizations of these likelihood maps over T_{eff} and A_V are shown in Bailer-Jones 2010b).

Both the degeneracy map and the likelihood map seem reasonable ways of showing the range of solutions possible for a given measurement, p. So which should we use to report the AP uncertainties? The degeneracy map is useful for showing which spectra we might observe for a given set of APs when we take into account the expected noise. However, the actual problem we face is determining the most likely APs given the measurement, so from the Bayesian perspective it is the likelihood map we should report to show the likely range of solutions. It represents the posterior PDF (with flat priors) and reports the 99% (or whatever we choose) most likely solutions for a given observation. The degeneracy map, in contrast, shows all possible APs which have spectra which agree with the measured data with a probability of more than 1% (or whatever we choose), regardless of the relative probabilities of these.

Furthermore, it is the likelihood map – and not the degeneracy map – which we should use when combining the outputs of ILIUM or any other parametrizer with other results. This is shown in Bailer-Jones (2010b) where the likelihood map is combined not with a flat prior but with a prior which optionally incorporates the measured parallax, the apparent magnitude the Hertzsprung–Russell Diagram and constraints on interstellar extinction.

5 Relation between the degeneracy map and the analytic approximation of the ILIUM uncertainties

How does the degeneracy map relate to the analytic approximation for the ILIUM errors given in section 5.2 of CBJ-046, equation 17? This relates the covariance in the APs to the covariance in the data and the sensitivity matrix

$$C_\phi = (S^T S)^{-1} S^T C_p S (S^T S)^{-1} \, . \tag{7}$$

This shows how the covariance in a single measured spectrum transforms to an equivalent covariance in the APs under the assumption that small changes in the spectra are related to small changes in the APs by $\delta p = S \delta \phi$. It only applies for small changes, and thus for small errors, because it involves a first order Taylor expansion, but it does not make any assumptions about the distribution δp or $\delta \phi$. If we now assume that the photometric errors are Gaussian with covariance C_p, then equation 4 gives the PDF for any observed spectrum, p, given the expected (forward model-predicted) spectrum p^0 for a given set of APs ϕ. Equation 7 transforms this into a Gaussian in the APs with covariance C_ϕ (and mean ϕ) which shows the equivalent information, but now as a function of the APs. Note that this does *not* make it into a posterior PDF over the APs: This Gaussian in the APs remains a likelihood of the data for a given AP solution, namely the one found by ILIUM. We may therefore expect that, in the limit of small errors, it would be the same as the degeneracy map, because this too identifies all spectra which
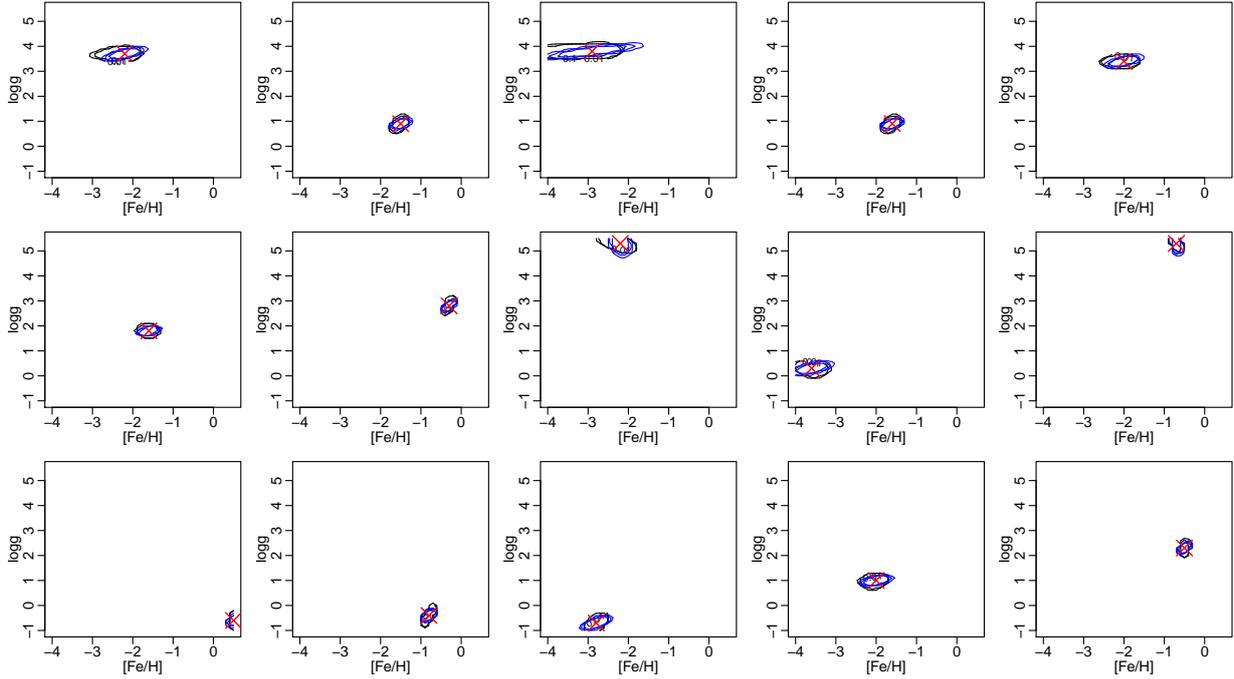
FIGURE 16: As Fig. 9 (but only retaining the contours at 0.1 and 0.01, which contain 10% and 99% of the probability respectively) showing in addition in blue the contours of the Gaussian approximation of the ILIUM errors at 1, 2 and 3 sigma from the peak.

are consistent with the observed spectrum to within some probability.

I investigate this relation empirically on for the $\log g$ and [Fe/H] degeneracy at fixed T_{eff} in the TGM problem (section 2.3, Fig. 9). We imagine to have estimated the APs of a spectrum as being the red cross shown in Fig. 9. The analytic approximation says that the AP uncertainties are represented by a Gaussian with this mean and the two-dimensional part of the covariance matrix C_ϕ for $\log g$ and [Fe/H]. Contours of this Gaussian, which are ellipses, are plotted in blue in Fig. 16. I have plotted contours at 0.607, 0.135 and 0.0111 times the peak of the Gaussian, which correspond to the 1, 2 and 3 sigma levels of a bivariate Gaussian. (Do not confuse these values with an integrated probability! Ideally I would show the contour which contains 99% of the integrated probability in order to be directly comparable with the 0.01 empirical contour, but the 3 sigma contour is close enough.) The degeneracy map contours are overplotted in black. These agree nicely with the analytic approximation in both size and orientation. In this particular case we may not want to refer to these APs as being "degenerate", although their errors are correlated. It is obvious from the complex shape of the degeneracy at G=18.5 (Fig. 8) that this approximation will no longer hold at fainter magnitudes.

This is a confirmation of the correctness of the analytic error approximation. Nonetheless, as discussed at the end of section 4, the correct Bayesian uncertainties to report are not these, but some summary of the likelihood map.

6 Likelihood maps for the TAG problem using the ILIUM forward model in an MCMC algorithm

In section 5 we saw how we can calculate the likelihood map by using the forward model to regularly sample a grid. For large numbers of APs this regular sampling would not be very efficient, however. The Markov Chain Monte Carlo (MCMC) method (e.g. MacKay 2003) may be more efficient and/or more accurate because it adapts to sample regions of high probability density more densely than regions of low probability density. After an initial burn-in period, this method (strictly a class of methods) performs a random walk to sample the PDF. Using a suitable density estimation algorithm we can then plot these samples as a continuous probability distribution. Furthermore, we can integrate over the samples to provide estimates of the mean APs, their standard deviations etc. (As the samples have been drawn from the PDF we do not need to include volume factor when making these calculations.)

To illustrate this I have used the Metropolis MCMC algorithm `MCMCmetrop1R{MCMCpack}` in R to build posterior PDFs for the TAG problem. I adopt simple flat priors over the APs: they are flat over a range slightly wider than the limits of the grids and zero outside: $P'(\phi) \propto 1/(\phi_{\max}-\phi_{\min})$, with ranges $\log g = [-1.0, 6.0]$, $\log(T_{\text{eff}}) = [\log(3000), \log(17000)]$, $A_V = [0, 12]$. Their product, suitably normalized, gives the prior $P'(\phi)$ in equation 3. The MCMC code works with the logarithm of the posterior

$$\ln[P'(\phi|\boldsymbol{p})] \;=\; -\frac{1}{2}D^2 - \frac{1}{2}\ln(2\pi) - \left(\sum_i \ln(\sigma_i)\right) + \ln[P'(\phi)] - \ln[P'(\boldsymbol{p})] \; . \tag{8}$$

Note that the prior must be correctly normalized so it has its correct size with respect to the one other term in this equation which varies with the APs (the first one). In contrast, $\log[P'(\boldsymbol{p})]$ and $\sum_i \ln(\sigma_i)$ depend only on the measured spectrum and does not distinguish between different AP solutions, so can be ignored.

As I have adopted flat priors over the APs, the resulting posterior PDF is entirely equivalent to the likelihood map, just produced by MCMC sampling rather than fixed regular sampling.

I applied this method to the validation set in the TAG problem at G=18.5 using a burn-in of 200 iterations followed by 1000 MCMC samples. (As it involves many more forward model evaluations it is of course considerably slower than ILIUM.) The initial APs were always taken as the mean of the training grid. Examples of the resulting PDFs over five stars are shown in Fig. 17. This experiment has used a relatively low number of iterations/samples in the Markov Chain, so these PDFs may not be very representative.

It is noteworthy that the posterior PDFs are very narrow, as were the likelihood maps, in the sense that the standard deviation of the PDF is narrow compared to the full range in the grid. In other words, the algorithm is very confident about its estimations (*high precision*). This is especially true of A_V which I think is fair to say is unreasonably precise. Such narrow PDFs are typical of the other cases, although of the 1000 stars in the validation set, the MCMC algorithm
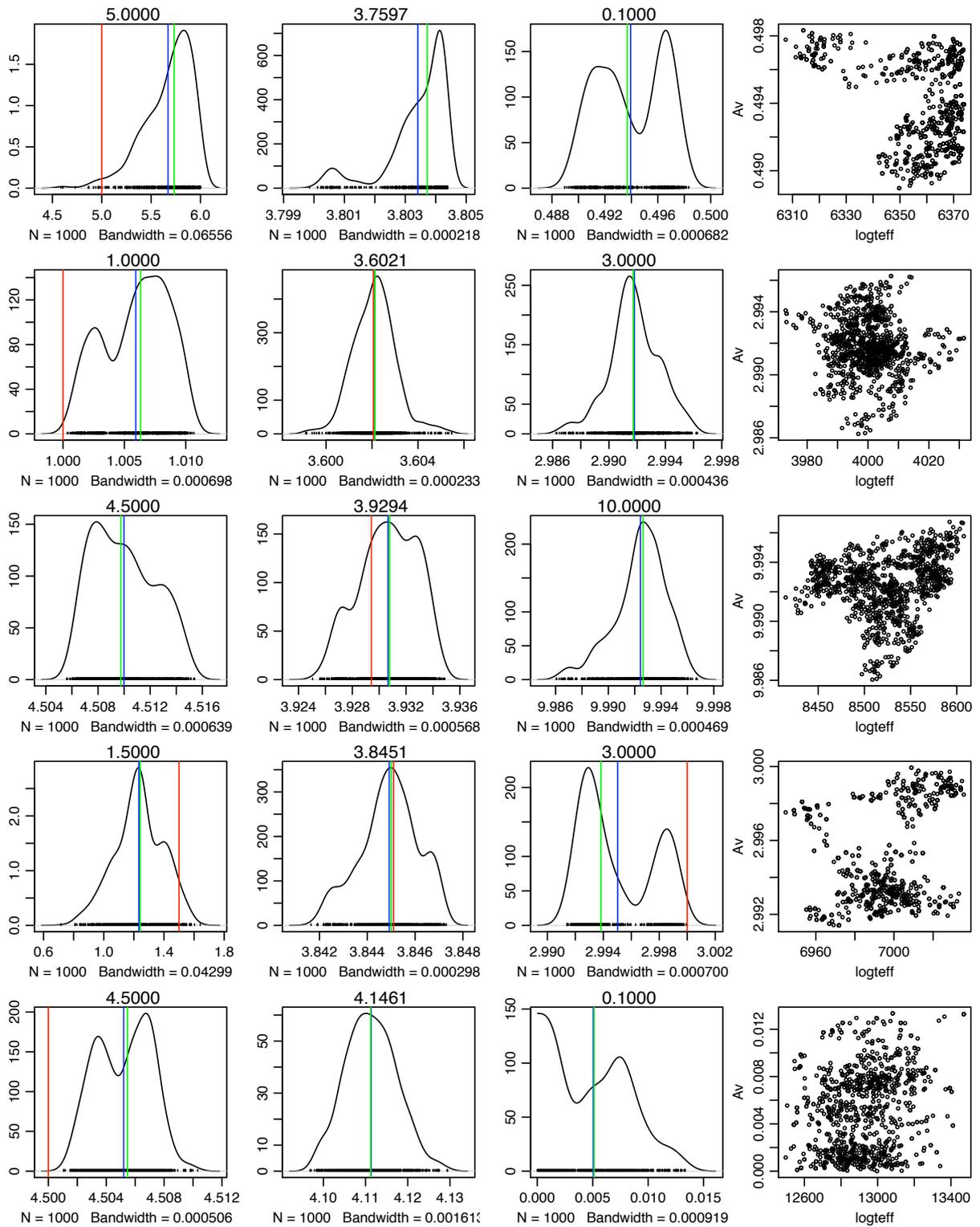
FIGURE 17: Posterior probability density function over $\log g$ (left column), $\log(T_{eff})$ (second column) and A_V (third column) for five stars (rows) calculated by by the MCMC/forward model applied to the TAG problem at G=18.5. The blue and green lines show the mean and median of the distributions; the red line is the true AP (also written at the top). The right column plots A_V vs. T_{eff} for the solutions.
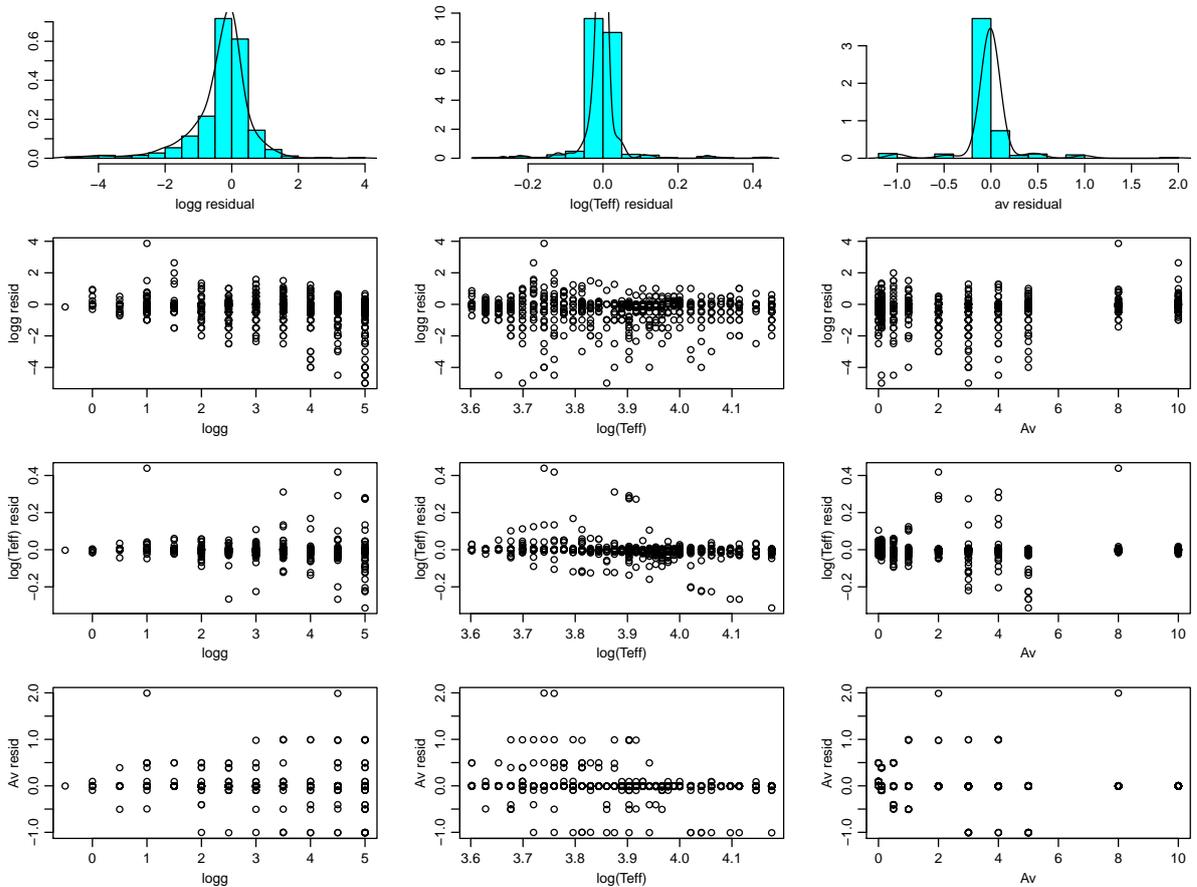
FIGURE 18: Residuals (estimated minus true) for the MCMC algorithm on the TAG problem at G=18.5 (only showing the 760/1000 stars where a solution was obtained)

failed on 240 of them (gave no solution). The reason for this is not yet clear (I have not tried that hard to optimize the algorithm), but it may well be the harder cases which failed.

If we integrate the PDF (take the mean of the samples), then we get an estimate of the APs. Fig. 18 plots these results by showing the distribution of the residuals (estimated minus true) and how the residuals vary with the true APs. We see a systematic in the $\log g$ residuals; a consequence of the weak $\log g$ signal (compared to the noise) at G=18.5. We also see that the A_V residual are partially discretized, which implies that the mean A_V estimates must also be for many objects (because the true values are also discrete). We do not expect this because the forward model and therefore the likelihood function is a smooth function of A_V. This needs to be explored further. Note that there is an apparent overdensity of residuals very close to zero for T_{eff} and A_V across the full range of these APs. Summarizing these results we get

Data plotted in rows, from top to bottom: all stars, Av<=1.0, Teff<=7000K
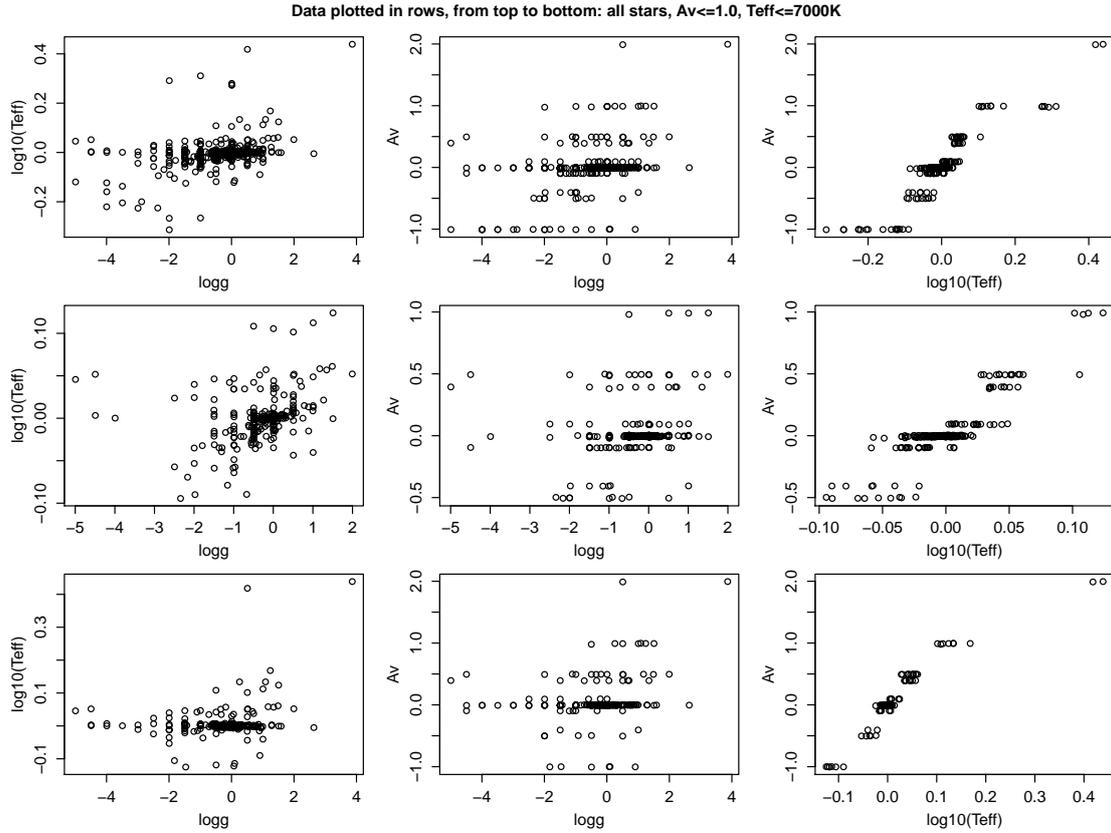
FIGURE 19: Correlation plots of the residuals (estimated minus true) for the MCMC algorithm on the TAG problem at G=18.5 (excluding the non-solutions), for the full data set (top row), low extinction stars (middle row) and cool stars (bottom row).

	$\log g$	$\log(T_{eff})$	A_V
$\overline{\delta\phi}$	-0.35	-0.0034	-0.0022
$\overline{|\delta\phi|}$	0.58	0.020	0.089
σ_ϕ	0.98	0.051	0.26

MCMC, TAG problem, G=18.5, excluding failures

where the statistics are the mean error (systematic), mean absolute error and RMS. These results are much better (*high accuracy*) than were obtained by ILIUM (cf. CBJ10a). However, it must be emphasised that MCMC failed to get any solution for 240/1000 stars, and these may well be the harder cases which give rise to larger errors. (In contrast ILIUM always gives a solution.) So these summary statistics should be interpreted with caution. Fig. 19 show the correlations between the residuals, where we again see the discretization of the A_V estimates. We also see a correlation between the T_{eff} and A_V residuals, which suggests that there is a degeneracy in the data, even though the MCMC samples of the likelihood function (right column of Fig. 17) have not revealed this. This suggests this MCMC application could be improved, probably by using a longer burn-in and/or more samples.

Repeating the experiment at G=15 we get the following

	$\log g$	$\log(\mathrm{T_{eff}})$	$\mathrm{A_V}$
$\overline{\delta\phi}$	-0.038	-0.00036	-0.0030
$\overline{|\delta\phi|}$	0.068	0.0017	0.0058
σ_ϕ	0.018	0.0027	0.008

MCMC, TAG problem, G=15.0, excluding failures

This too is better than ILIUM, but here MCMC gave no solution for 333 of 1000 objects.

7 The ILIUM training philosophy and incorporation of other data

As noted in CBJ10a, ILIUM requires a grid regular in the strong APs. This is a prerequisite for the independent forward modelling of the strong and weak APs and is easily satisfied by synthetic atmospheric grids. However, it is virtually impossible to satisfy with synthetic grids built from primary APs: these are grids where we specify mass, age and composition of stars and then derive $\mathrm{T_{eff}}$ (and $\log g$) via interpolation. (This is done to create the "random" grids for some of the CU8 simulations, as described in Sordo & Vallenari 2008 and Bailer-Jones 2007). However, this should not be seen as a genuine limitation, a key principle of ILIUM is to obviate the need to use interpolation to define the training set. The ILIUM philosophy is to allow us to work directly from the grid and do the interpolation as part of the AP estimation only and as needed. In contrast, a standard (inverse) modelling method (e.g. an SVM) which is trained on data derived from a primary grid effectively interpolates this grid a second time (as an inverse function) as part of its training process.

ILIUM's restriction to grids regular in the strong APs also means it could not be trained directly on empirical grids. But training on empirical grids is again not consistent with the ILIUM philosophy, because such empirical grids would themselves have to be parametrized using some other method, resulting in a complete "double" parametrization. Instead the idea is to calibrate a regular synthetic atmospheric grid using a sparse grid of classified empirical data. A procedure for doing this was outlined in Bailer-Jones (2009e).

ILIUM itself is not a probabilistic method, but we have seen in this technical note how the ILIUM forward model can be used to define a complete Bayesian posterior probability distribution over the APs. This idea is extended in Bailer-Jones (2010b) to permit use the relevant information on stellar parameters provided by parallax and apparent magnitude, as well as prior information based on the Hertzsprung–Russell Diagram. In principle the ILIUM input space and training grid could be expanded to include other data, such as parallax or the RVS spectrum. However, this probabilistic method gives a more elegant way of combining independent AP estimations from different data sources. Bailer-Jones & Smith (2010) describe more generally how to combine parametrization models based on different data.

References

Bailer-Jones C.A.L., 2007, *CU8 Cycle 3 simulations requirements*, GAIA-C8-SP-MPIA-CBJ-029

Bailer-Jones C.A.L., 2009a, ILIUM*: An iterative local interpolation method for parameter estimation*, GAIA-C8-TN-MPIA-CBJ-042

Bailer-Jones C.A.L., 2009b, *Application of* ILIUM *to the estimation of the* T_{eff}*–[Fe/H] pair from BP/RP*, GAIA-C8-TN-MPIA-CBJ-043

Bailer-Jones C.A.L., 2009c, ILIUM III. *Further observations, tests and developments*, GAIA-C8-TN-MPIA-CBJ-046

Bailer-Jones C.A.L., 2009d, ILIUM IV. *Three-dimensional forward model and demonstration of a strong and ubiquitous* T_{eff}*-*A_V *degeneracy in BP/RP* GAIA-C8-TN-MPIA-CBJ-048

Bailer-Jones C.A.L., 2009e, *A procedure to calibrate algorithms for estimating parameters from spectra*, GAIA-C8-TN-MPIA-CBJ-044

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 (CBJ10a)

Bailer-Jones C.A.L., 2010b, *Probabilistic combination of AP estimates based on spectra, astrometry and the HR Diagram with the aim of reducing degeneracy*, GAIA-C8-TN-MPIA-CBJ-049

Bailer-Jones C.A.L., Smith K.S., 2010, *Combining probabilities*, GAIA-C8-TN-MPIA-CBJ-053

Knude J., Lindstroem H., 2007, *Interstellar extinction estimation. Using color-ratios from BP, RP, G,* G_{RVS} *and JHK photometry*, GAIA-C8-TN-NBI-JK-002-1

MacKay D.J.C, 2003, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, available online

Sordo R, Vallenari A., 2008, *Description of CU8 cycle 3 simulated data*, GAIA-C8-DA-OAPD-RS-002

Tsalmantza P., Bailer-Jones C.A.L., *Semi-empirical library of stellar spectra from SDSS*, GAIA-C8-TN-MPIA-PAT-004