A procedure to calibrate algorithms for estimating parameters from spectra

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-044
issue: 1
revision: 0
date: 2009-02-07
status: Issued

Abstract

I outline a procedure for calibrating the Gaia spectral parameter estimation algorithms, in particular GSP-Phot. It assumes that synthetic spectra accurately reproduce small changes in the APs. In contrast, they may not accurately reflect large changes in the APs or they may have systematic offsets in the absolute fluxes. It is these changes that the outlined calibration procedure can correct for, using Gaia observations of stars with known APs (*AP reference stars*). The method works by removing from the synthetic data the low frequency variation of flux with AP and replacing it with that determined from the calibration data. It is based on the forward modelling approach discussed in GAIA-C8-TN-MPIA-CBJ-042. To apply in practice would require several hundred stars covering a range of $\mathrm{T_{eff}}$, $\log g$ and [Fe/H]. Calibration of the interstellar extinction remains an open issue.

Document History

Issue	Revision	Date	Author	Comment
1	0	2009-02-078	CBJ	First issue, incorporating comments from Caroline Soubiran and Andreas Korn on section 3.
D	0	2009-01-28	CBJ	First draft

Contents

1 Introduction

Many algorithms for estimating astrophysical parameters (APs) from spectra are trained directly on synthetic spectra. Such spectra have the advantage that we can generate them at any resolution, over any wavelength range, for any combination of APs, and for any level of SNR (including zero), permitting us to construct almost any grid which we desire. By varying the input parameters (e.g. the stellar structure or atmosphere parameters in the case of stars), we can investigate the impact of physical parameters on the spectra.

Synthetic spectra have the crucial drawback, however, that they are not exact replicas of real stars. (I'll discuss just stars from now on for simplicity.) Real stars show cosmic variance which we cannot explain or have difficulties modelling. This may be due to uncertain data (line opacities, abundance ratios), uncertain physics (diffusion, convective overshoot) or effects of parameters which, for whatever reason, we may choose to ignore (e.g. rotation or variations in the extinction law). I call this the *synthetic spectra mismatch problem*. This expresses itself in a number of ways. A useful and relevant way to think about it is to consider the variation of the flux in some band with respect to one AP, say $\mathrm{T_{eff}}$, holding the others fixed. The variation may not be the same for synthetic data as for real data.

Note that I use the term *band* to refer to any sampling of the spectral energy distribution. So a band could be a photometric filter, a pixel in a spectrum, a sum of several pixels, or an integrated line-spread function over the spectrum.

The main consequence of the synthetic spectra mismatch problem is that when classifying real spectra using synthetic spectra (or rather, a model trained on synthetic spectra), we are likely to make errors. These error are systematic in the sense that – in the absence of photometric noise – any given star will always be assigned APs which differ from its true APs by a fixed amount.

Here I describe how Gaia observations of stars with known APs, which I call *AP reference stars*, may be used to correct for the synthetic spectra mismatch problem.

2 Calibration procedure

The basic idea is to use Gaia observations of the AP reference stars to modify the fluxes of all objects in the synthetic grid. This produces a hybrid synthetic–real grid which is then used for training. The procedure is based on the concept of forward modelling, described in more detail in GAIA-C8-TN-MPIA-CBJ-042 (Bailer-Jones 2008).

2.1 Assumptions

There are four fundamental assumptions behind the procedure.

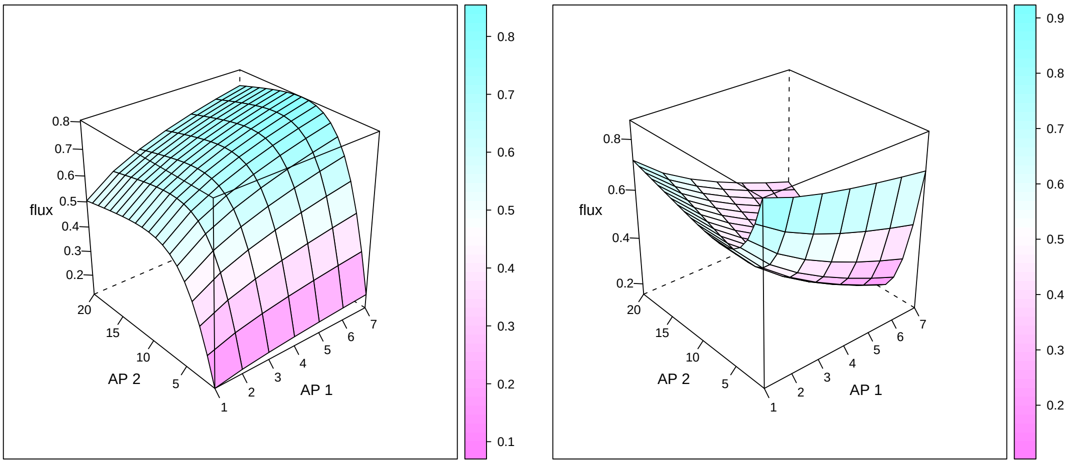

FIGURE 1: Schematic of generative functions over two APs. Each panel shows the function for a different band

Assumption 1 There exists a unique relationship between the flux, p, in a band and the APs, $\boldsymbol{\phi}$, which I call the *generative function*, $g(\boldsymbol{\phi})$. There is a separate function for each band, i, so $p_i = g_i(\boldsymbol{\phi}) + \epsilon_i$, where ϵ_i is a noise term.

Assumption 2 The generative function is *smooth*. This is a relative term, but it assumes that at some length scale in $\boldsymbol{\phi}$, samples from the generative function can be accurately interpolated to reproduce the generative function to arbitrary accuracy. A lower limit on the length scale for $\mathrm{T}_{\mathrm{eff}}$ is 10 K and for [Fe/H] and $\log g$ is 0.1 dex.

Note that in general $\boldsymbol{\phi}$ is a vector, i.e. the generative function is a multidimensional function. Schematic illustrations of such functions for two APs, for two different bands, are shown in Fig. 1.

We do not explicitly know the generative function, i.e. we don't have an analytical formula for it, so we can only sample it using our stellar atmosphere models. If we sample it at sufficiently high AP resolution and then fit this, we get the so-called *forward model*, which we use as a proxy for the generative function. (This uses Assumptions 1 & 2.) The distinction is not that important in this technical note – we can consider the forward model to fit the generative function exactly – but in other contexts it is (e.g. in CBJ-042 when we are concerned with fitting "strong" and "weak" APs). For consistency with other work, I will now refer to the forward model.

Assumption 3 Stellar models accurately reproduce the relative fluxes in a band for small changes in the APs.

That is, if we change the $\mathrm{T_{eff}}$ by, say, 1%, then the change in the true flux is accurately replicated by the synthetic spectra. The absolute values of the flux and the flux variations across larger AP ranges, however, are not assumed to be accurately reproduced by the stellar models.

Assumption 4 We can obtain accurate APs of a set of stars which Gaia will observe, covering a broad range of the APs of interest. These are the *AP reference stars*. The AP sampling is sufficient to map the large scale variation of flux with APs.

"Accurate" here means with errors on the order of (or less than) that which we hope to achieve with Gaia. The issue of the number and AP sampling of these will be discussed later. In principle Gaia would not have to observe them: we could take higher spectral resolution spectrophotomtetry and simulate the Gaia bands using a simulator (e.g. GOG). This permits us to correct for the synthetic spectra mismatch problem. However, it would not allow us to correct for a second, independent, problem, namely that GOG does not and cannot accurately simulate the instruments. By using Gaia observations of the AP reference stars, we can correct for this too.

2.2 Procedure

The procedure is simple and is described with the help of Fig. 2. I first consider variation with respect to just one AP, $\mathrm{T_{eff}}$. The figure shows (in black) the forward model as a function of $\mathrm{T_{eff}}$ for four BP/RP bands constructed from the Cycle 3 GOG simulations for the main stellar library (Sordo & Vallenari 2008; see also section 2 of CBJ-043). ($\mathrm{A_V}$ and [Fe/H] are zero and the forward model fit marginalizes over a small range of $\log g$).

Pretend data on some AP reference stars are shown as blue points. Given Assumption 4, a low order fit to these, L_y, (solid blue line) reproduces the large scale variation of flux with $\mathrm{T_{eff}}$. We remove the overall shape of the forward model, f, by dividing by its own low order fit, L_f (green dashed line). We then multiply the result by the low order fit to the calibration data. This gives the *calibrated forward model*, $f^{'}$, shown in red

$$f^{'} = \frac{f}{L_f} \times L_y$$

That is, we have replaced the large scale variation of the forward model (which may not be accurate) with the large scale variation of the calibration data.

In three of the four panels, 362 nm, 573 nm and 857 nm, I have made both low order fits using a smoothing spline with 5 degrees of freedom. In these cases we see that the calibrated forward model retains its original high frequency structure, but has adopted the overall shape and flux scale of the calibration data. This is about the level of smoothing one would expect to perform, although the mismatch between the calibration data and the synthetic data is only for illustration

TABLE 1: Function/data definitions

function/data	symbol	how shown in plots
forward model	f	black solid line
low order fit to forward model	L_f	green dashed line
calibration data	y	blue points
low order for to calibration data	L_y	blue solid line
calibrated forward model	$f^{'}$	red solid line

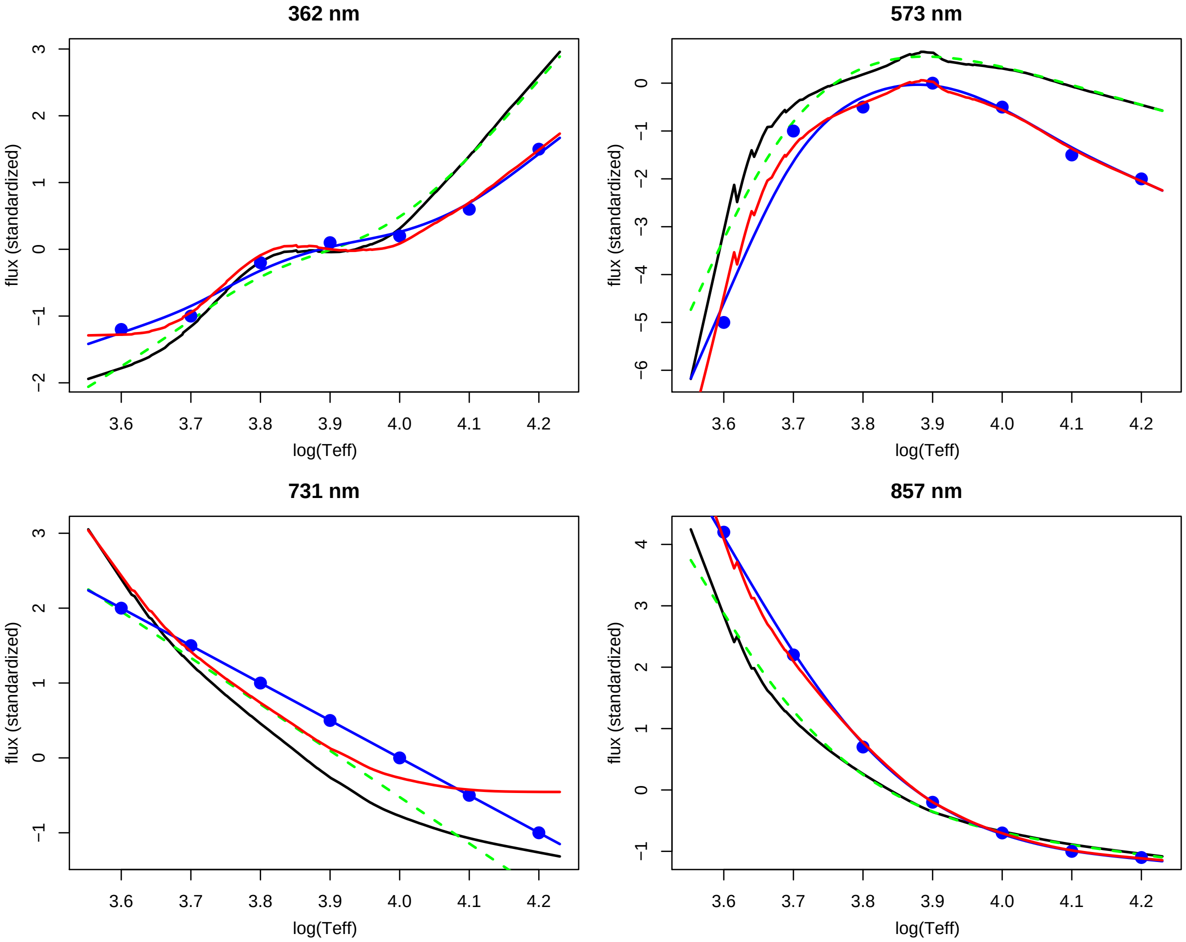

FIGURE 2: Forward model calibration procedure for four bands (central wavelength given at the top of each panel). red $= \frac{\text{black}}{\text{green}} \times$ blue . See Table 1 for the definitions.

purposes and not meant to be representative. In the bottom left panel, for 731 nm, I examine what happens if I use a lower order (linear) fit. This may not be desirable, but it is a useful exercise for identifying the origin of the structure in the calibrated forward model.

Calibration may, in practice, be much simpler for some bands, e.g. it may just involve a constant offset. On the other hand, if we had sufficient calibration data to get an accurate fit to L_y over the full AP range, then we would not need to use synthetic data at all for training. The data on the AP reference stars could then act as the grid to which we fit a calibrated forward model directly.

It's worth reiterating one of the assumptions: The procedure requires that the ratio of fluxes in bands (spectral regions) between two stars with small differences in their astrophysical parameters are accurately reproduced by the synthetic spectra. This is necessary so that we can use low order fits (smoother than the forward model itself). But it does *not* assume that the two spectra differ by a single factor, common to all bands. Each band is corrected independently.

2.3 Dealing with multiple APs

So far I have only looked at the problem of calibrating the fluxes with variation in one AP. In principle we can use the same approach but with the forward model and calibration data and their fits extended to multiple dimensions, e.g. a 2D fit to $\mathrm{T_{eff}}$–[Fe/H] surfaces like those illustrated in Fig. 1. My experience with BP/RP data is that [Fe/H] and $\log g$ vary just as smoothly with flux as $\mathrm{T_{eff}}$ does, or even more so. However, fitting such multi-dimensional models is complicated in practice by the presence of both "strong" APs and "weak" APs, at least for BP/RP. These are defined as APs which have a large and small (respectively) impact on the variance in the fluxes, such as $\mathrm{T_{eff}}$ and $\mathrm{A_V}$ in the former category and [Fe/H] and $\log g$ in the latter. This problem was discussed in CBJ-042 in the context of developing a parameter estimation algorithm (ILIUM) designed to cope with this. We can follow a similar approach here, independently fitting 1D forward models to the APs either marginalizing over the weak APs (for the fit against the strong APs) or at fixed values of the strong APs (for the fit against the weak APs). See CBJ-042 for more details. One complication is that the method outlined there requires a "semi-regular grid", viz. one consisting of spectra with a range of values of the weak APs for each combination of the strong APs. This cannot be fulfilled by the AP reference stars, but is probably an issue in the method we can overcome.

3 Which calibration data do we need?

The amount of calibration data required for this procedure depends on (1) how much synthetic data deviate from real data, (2) how smooth is the variation of the flux with APs, (3) the spectral resolution (i.e. BP/RP or RVS – I consider here BP/RP). To first order, the sampling of the APs used by the stellar libraries for the cycle 3 data represents an upper limit to the sampling (and thus number of objects) required (although this *may* be because I haven't plotted data sampled at higher AP resolution to know whether this is really enough!). For the full AP range of the cycle 3 data (M to O stars, dwarfs to supergiants, [Fe/H] from -4.0 to $+1.0$) this is 4364 stars (Sordo & Vallenari 2008). (2343 Marcs, 2018 Basel, 203 OB library – the OB library is just for [Fe/H] = 0.0). We certainly would not need as many stars, because the sampling in $\mathrm{T_{eff}}$ for OBA stars is higher than needed, as is the sampling of metallicity around solar metallicity (see Fig. 1 of CBJ-043). Probably only half the density is required in each of $\mathrm{T_{eff}}$ and [Fe/H], especially when we consider that we'll only need to extend [Fe/H] down to -2.5 or -3.0. (Gaia will probably see many stars with even lower metallicities, and GSP-Phot might just be able to identify the [see CBJ-043]. But we don't know enough of them at the moment to be able to include that many in the AP reference star grid.) We could likewise get away with an average $\log g$ step size for the AP reference stars larger than the 0.5 dex used in the synthetic grids. Overall, we probably need only $1/2^3 = 1/8$ as many objects, some 550 stars.

GSP-Phot will additionally attempt to estimate [α/Fe], $\mathrm{A_V}$ and $\mathrm{R_V}$. [α/Fe] is only relevant for cool stars. It shows some correlation with [Fe/H] in the Galactic population, but with a large scatter. Allowing for variation in this perhaps only adds another hundred or so stars. Extinction is a difficult parameter (or set of parameters) and is probably best not calibrated using real data, not least because of the difficulty in estimating it observationally. This implies that all AP reference stars should have zero extinction, a condition which will be hard or impossible to achieve. Therefore it is likely that we would be forced to calibrate for the low levels of extinction found in the reference stars. This issue needs to be examined further.

The data needed for the AP reference stars must be sufficient to estimate the APs with a precision similar to (but preferably better than) that which we can achieve with the AP estimation algorithms (some estimates of GSP-Phot performance can be seen in CBJ-042, -043 and CT-006). This would presumably be medium or high resolution optical spectra (extending from the UV atmospheric cutoff to the red). The APs could be estimated via "conventional" techniques of line analysis or via machine learning methods. Whatever method we use depends ultimately on a set of stellar models, but this is a prerequisite of any *physical* parametrization system. Indeed, at a fundamental level, our physical descriptions of the stars in terms of effective temperature, surface gravity, abundance etc. are a limited description of real stars and observables.

The AP reference stars should be sufficiently bright so that they can be observed with high SNR by Gaia, ideally limited only by the flux calibration floor of Gaia, which I I take to be 0.3% per non-oversampled BP/RP pixel. A SNR of this level is achieved in the end-of-mission data for most types of stars across much of the spectrum at around G=15 (although this needs to be studied much more carefully). It would be desirable get high SNR data on the reference stars earlier in the mission, perhaps even with single epoch spectra (which would also circumvent possible problems with spectral combination). This requires the stars to be around 2.3 mag brighter, say G$\simeq$13, assuming that gates are not yet used to limit the integration time at this magnitude (the current plan is only to use gates at brighter magnitudes). Indeed, to avoid possible flux calibration problems with the gates, we *may* want to avoid having any AP reference stars so bright that they are observed in this mode (although I do not see that the gates should present any particular problem). It may also be that some AP combinations are only known for stars fainter than G=15. In that case, we may want to populate that part of the AP reference star grid at a higher density in order to compensate for the lower SNR. Note that there is no requirement to have the AP reference stars spread over the whole sky. We do need them to be sufficiently isolated (uncrowded regions) so that we can obtain unambiguous Gaia and ground-based data.

Heiter et al. (2008) have suggested the requirements on the data on the AP reference stars needed to calibrate the General Stellar Parametrizer algorithms (GSP-Phot and GSP-Spec) in Gaia. They present the AP ranges and number of objects needed to sufficiently sample the variations in $\mathrm{T}_{\mathrm{eff}}$, $\log g$ and [Fe/H] in the stellar populations Gaia will see (which is basically all of them) and have identified spectral/AP catalogues which could be used for this purpose. They propose three levels of reference stars – according to the AP accuracy and thus data quality required – which bootstrap from one level to the next. The middle one of these, which they call the primary grid, is the closest in terms of sampling and requirements to what I outline above. They suggest a near-uniform grid in $\mathrm{T}_{\mathrm{eff}}$, $\log g$ and [Fe/H] with step sizes of 7%, 1.0 dex and 0.5 dex respectively which would comprise some 560 stars spanning the range $4000\,\mathrm{K} \leq \mathrm{T}_{\mathrm{eff}} \leq 7000\,\mathrm{K}$, $-0.5\,\mathrm{dex} \leq \log g \leq +5.0\,\mathrm{dex}$ and $-3.0\,\mathrm{dex} \leq \mathrm{[Fe/H]} \leq +0.5\,\mathrm{K}$ (their Table 1). This is actually denser than what I estimated above, because they only consider FGK stars, but the OBA stars probably only add a couple of hundred more calibrators, so the final numbers are similar. Heiter et al. have already started to identify catalogues and open and globular clusters for the selection of primary grid stars, as well as targets and facilities for new observations.

References

Bailer-Jones C.A.L., 2008, ILIUM*: An iterative local interpolation method for parameter estimation*, GAIA-C8-TN-MPIA-CBJ-042

Bailer-Jones C.A.L., 2008, *Application of* ILIUM *to the estimation of the* $\mathrm{T}_{\mathrm{eff}}$ *– [Fe/H] pair from BP/RP*, GAIA-C8-TN-MPIA-CBJ-043

Heiter U., Soubiran C., Korn A., 2008, *Calibration of the General Stellar Parametrizer algorithms using ground-based observations*, GAIA-C8-TN-UAO-UH-001-1

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

Tiede C., 2008, *GWP-S-822 GSP-Phot. Cycle 4 Software Test Plan and Verification Report*, GAIA-C8-SP-MPIA-CT-006-4