Proc. 28th Ann. Conf. GfKl2004, submitted

Design of astronomical filter systems for stellar classification using evolutionary algorithms

C.A.L. Bailer-Jones

Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany

Abstract. I present a novel method for designing filter systems for astrophysical surveys. The filter system is designed to optimally sample a stellar spectrum such that its astrophysical parameters (APs: temperature, chemical composition etc.) can be determined using supervised classification methods. The design problem is addressed by casting it as an optimization problem: A figure-of-merit (FoM) is constructed which measures the ability of the filter system to vectorially 'separate' stars with different APs; this FoM is then optimized with respect to the parameters of the filter system using an evolutionary algorithm. The resulting filter systems are found to be competitive in performance with conventionally designed systems.

1 Astrophysical context

Astrophysics relies on large statistical surveys of astronomical objects for advancing our understanding of the cosmos. In stellar astrophysics, for example, by measuring the spectral energy distribution (spectrum) of many different types of stars across our Galaxy we can gain insight into the formation and evolution of stars and of the Galaxy itself. Ideally we would obtain high quality spectra of literally billions of stars, from which we can determine stellar intrinsic properties, or *astrophysical parameters (APs)*, quantities such as the temperature, chemical composition and surface gravity. However, for various technical reasons such detailed spectroscopy is not (yet) possible. Instead, we must limit ourselves to photometry, that is, coarsely sampling a spectrum at pre-defined locations with a filter system (for an example see Fig. 4). By analysing the spectrum of a specific star in detail, we could design a filter system which is adequate for determining the APs of that type of star to some desired accuracy. However, large surveys must observe *many* different types of stars with a *single* filter system. Hence this filter system must be some kind of optimal average system, the design of which is furthermore subject to numerous instrumental constraints.

A number of upcoming surveys are therefore faced with the difficult question of how to define their optimal filter system. Existing systems have been designed for more specific purposes or for more restricted classes of objects than is appropriate for these new surveys. The 'conventional' approach to this problem is to manually modify existing systems based on the best of our astrophysical knowledge. Yet given the numerous conflicting requirements

placed on the filter system, this is unlikely to be very efficient or even successful. Moreover, one would still not known whether a better filter system could be constructed within the constraints.

In this paper I outline a more systematic approach to the filter design problem called Heuristic Filter Design (HFD). It is assumed that we have a set of stellar spectra (the *grid*) which represents the kind of stars that will be present in the survey. Their spectra and their APs we want to determine are known. Each of these APs varies continuously across the grid. The survey instrument (which defines the size of mirror, type of detectors etc.) is also fixed and represents the constraints within which the filter system is optimized: given a set of filters, this instrument model allows us to calculate the amount of light (number of photons per unit wavelength) measured for each star in each filter, plus their expected errors. In some sense, filter design boils down to the optimal partition of these photons into different filters. HFD is being used to aid the design of the filter system for a major future astronomical survey (see my other contribution in these proceedings for one major survey). More details on the model and its application can be found in Bailer-Jones (2004).

This filter design problem is closely related to the problem of determining the J APs of a star from measurements in I filters. This latter problem is usually solved using supervised multivariate regression methods, that is, given a set of pre-classified filter data we apply a regression method (such as a neural network or minimum distance algorithm) to establish the data–AP mapping (Bailer-Jones 2002). HFD can be seen as a partial inversion of this problem in which we essentially optimize the data space itself in order to simplify its topology with respect to the APs. This should increase the performance of an ideal regression model fitted to these data and/or permit a simpler model.

2 The optimization model

2.1 Parametrization

A filter is parametrized with three parameters: the central wavelength, c, the half-width at half maximum (HWHM), b, and the fractional integration (or exposure) time, t, i.e. the fraction of the total integration time available per star which is allocated to this filter. (The instrument model specifies the total time available per star.) The profile a filter – the fraction of light transmitted at each wavelength, λ – is given by the generalized Gaussian

$$\Psi(\lambda) = \Psi_0 \exp\left[-(\ln 2)\left|\frac{\lambda - c}{b}\right|^{\gamma}\right] \quad . \tag{1}$$

This is Gaussian for $\gamma = 2$, and rectangular for $\gamma = \infty$. $\gamma = 8$ and $\Psi_0 = 0.9$ are used. For a system of I filters there are therefore $3I$ parameters which

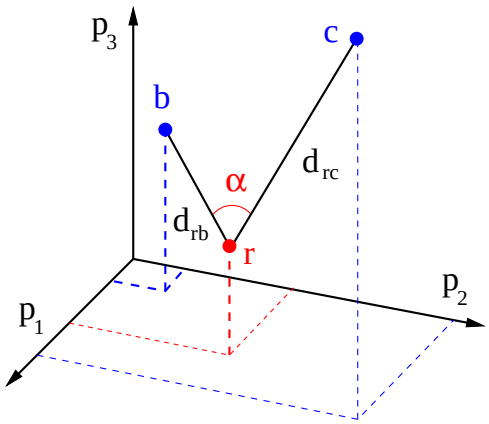

Fig. 1. A three-dimensional data (filter) space: p_i is the number of photons collected in filter i. For star r in the grid we identify its two nearest neighbours (assuming the number of APs is two), b and c, each of which differs from r in just one of the APs. The scalar distance to these neighbours defines the AP-gradient and the angle between their vectors, $\sin\alpha$, the 'vector separation'. An optimal filter system (for r) has $\alpha = 90°$ and the AP-gradients large.

must be optimized. The optimization is performed within practical limits (set by the instrument): c and b are limited such that no part of any filter has a significant transmission (Ψ) outside of the wavelength range 2750–11250 Å. Additionally, the maximum HWHM is restricted to about 4000 Å. t must of course be $0.0 \leq t_i \leq 1.0$ and be normalized, $\sum_i t_i = 1.0$ (i labels a filter).

2.2 Figure-of-merit (fitness)

The I filters of a filter system define an I dimensional data space in which the measured objects (stars) reside (see Fig. 1). The purpose of the filter system is to allow us to use these data to determine the J APs per stars. At any point in this space, each AP will vary in a certain direction (the *principal direction*), and at a certain rate, the (scalar) *AP-gradient*. These we can calculate, or at least approximate, using the grid of pre-classified stars. In order to be able to determine the J APs, we clearly need $I \geq J$, but we must also ensure (1) that the AP-gradient is sufficiently large so that, given the signal-to-noise ratio (SNR) in the data, we can determine the AP to the desired precision, and (2) that the principal directions for each AP are mutually orthogonal, or as close to this as possible (otherwise the APs are partially degenerate). In other words, the goal of the filter system is to maximally 'separate' the different APs for the different stars in a vectorial sense.

These ideas are converted into a figure-or-merit, or fitness, as follows. For each star, r, in the grid, we find its J nearest neighbours, each of which differs from r in only one of the J APs. The relevant 'distance' between r and that

neighbour differing in AP j (call it n_j), is the *AP-gradient* and is defined as

$$h_{r,n_j} = \frac{d_{r,n_j}}{|\Delta\phi_{r,n_j}|} \tag{2}$$

where d_{r,n_j} is the Euclidean distance between r and n_j in SNR units and $\Delta\phi_{r,n_j}$ is their difference in AP j. Clearly, the larger h the better we have separated r and n_j. However, we must also minimise the degeneracy between the principal directions to these J neighbours, in other words, we want angle α in Fig. 1 to be as close to 90° as possible for all neighbour pairs. Combining these measures, we see that a useful figure-of-merit of separation is

$$x_{r,j,j'} = h_{r,n_j}\, h_{r,n_{j'}} \sin\alpha_{r,j,j'} \tag{3}$$

where j and j' label those neighbours which differ from r in APs j and j' respectively. Note that the above is simply the magnitude of the cross product between the two vectors. For J APs we have $J(J-1)/2$ pairs of neighbours and thus $J(J-1)/2$ terms like eqn 3. Summing these over all stars in the grid gives the final fitness which is to be maximized

$$F_k = \sum_{j,j'\neq j} \sum_{r} x_{k,r,j,j'} \quad . \tag{4}$$

(The actual fitness function is a slight modification which weights and transforms some of the terms to increase its sensitivity: see Bailer-Jones (2004)).

2.3 Evolutionary algorithm

An evolutionary algorithm (EA) is used for the optimization (e.g. Bäck & Schwefel (1993)). A population of 200 individuals is evolved over 200 generations. An outline of the algorithm is shown in Fig. 2. Natural selection is emulated using the 'roulette wheel' method, i.e. objects are selected with a probability directly proportional to their fitness (eqn. 4). Elitism is used, meaning that the E fittest individuals are always selected (and are still subject to probabilistic selection). In common with many other EA applications, this is found to improve performance. $E = 10$ is used in the results shown, although $E = 50$ actually ensures more consistent convergence (independence of initial conditions). The two search operators are recombination and mutation. Recombination involves swapping a randomly chosen filter between two individuals. Mutation is implemented by adding a Gaussian random variable, $N(0, \sigma_c)$, to the central wavelength, c, and multiplying the HWHM, b, and fractional integration time, t, by $N(1, \sigma_h)$ and $N(1, \sigma_t)$ respectively. If a mutation would take a filter parameter out of bounds, then the mutation is rejected and that parameter passed on unchanged. The standard deviations σ_c, σ_h and σ_t were 500 Å, 0.5 and 0.25 respectively, and the mutation probability per parameter was 0.4. It was found that HFD was quite insensitive to

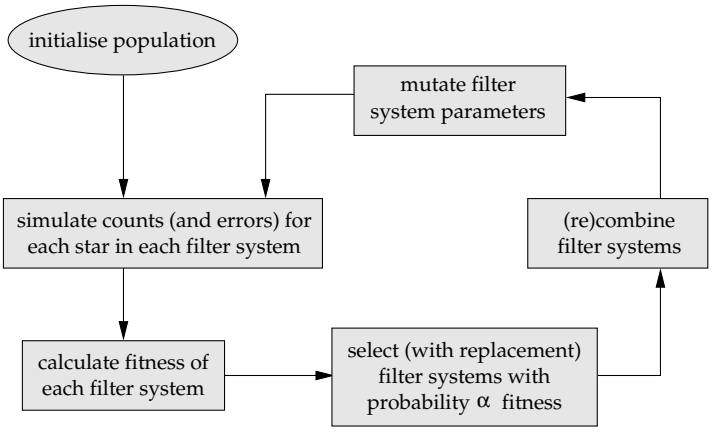

Fig. 2. Flow chart of the core aspects of the HFD optimization algorithm. A single loop represents a single iteration, i.e. the production of one new generation of filter systems.

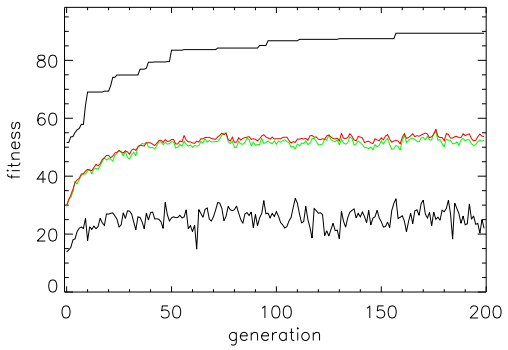

Fig. 3. Evolution of fitness statistics for a typical HFD run. The lines from top to bottom denote the maximum, mean, median and minimum fitness in the population.

the mutation probability (unless a very low probability is used, in which case there is rapid convergence to a poor local maximum) and to the standard deviations. The absence of recombination also made negligible impact.

3 Application, results and interpretation

HFD is applied to the design of a 10-filter system for determining four APs. The evolution of the fitness is shown in Fig. 3. The optimization was terminated after 200 iterations. Continuing for ten times as many iterations or using a population ten times as large produced no significant difference. The entire optimization was repeated 20 times from different initial (random) populations. The fittest filter system produced from this is shown in Fig. 4.

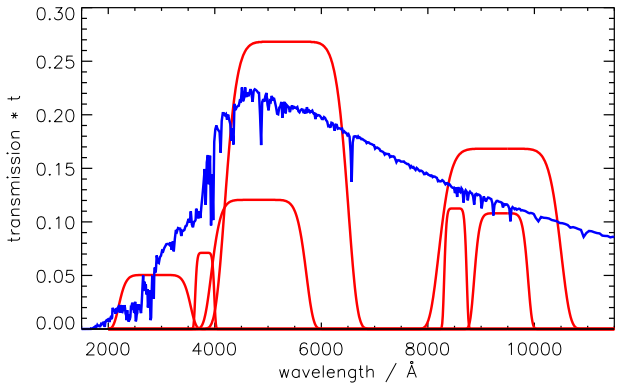

Fig. 4. An optimized filter system produced by HFD. Each of the seven filters is the plot of $\Psi(\lambda)$ from eqn. 1, but multiplied by the fractional integration time, t. Overplotted is an example of a stellar spectrum (number of photons vs. wavelength) arbitrarily scaled.

Inspection of the filter system shows that it consists of only seven filters, i.e. the optimization has 'turned off' three filters by setting their fractional integration times to zero. This is a recurrent feature. At low SNR it makes sense, because there is a penalty to be paid for retaining more filters (as the integration time must then be divided among more filters). It is also interesting that the system has naturally self-regulated the widths, b, of the filters. That is, their range is much narrower than the range permitted by the limits of the optimization. This is encouraging, as on pure SNR grounds wider filters are better as they collect more photons (Poisson statistics). Yet beyond a certain width this is detrimental to the separation of the APs. The fact that the central wavelengths, c, cover the whole permitted wavelength range is expected from what we know about stellar spectra: a wide coverage gives a good 'leverage' for determining small changes in the slope of the spectrum. However, the lack of filter coverage between 7000 and 8000 Å is curious, as is the fact that the wide filter between and 8000 Å 10 500 Å is almost equal to the sum of the two filters covering the same range. It could be that this is measuring small differences between the filters.

Astrophysically these filter systems are quite unconventional in two important respects.

First, the filters are very broad compared to filters typically used for stellar parameter estimation. Narrow filters are able to isolate individual spectral features that we know are sensitive to specific APs. Certainly, in an ideal case, such narrow filters could better isolate specific signatures. But this implicitly assumes that we only have to deal with a narrow range of stellar types so that we could employ such specific filters. In contrast, HFD has been applied

to a very broad grid of stars, as demanded by the planned surveys. Moreover, it has been applied to stellar parameters which can be demonstrated to have a broad band impact on the stellar spectrum (i.e. cause a variation which is coherent over a large wavelength range). In this case, broad band filters may be more efficient.

Second, the filters overlap in the wavelength domain. This is sometimes avoided, as it complicates the interpretation of plots of colour indices (a colour index is the ratio of the flux obtained in two filters). However, modern surveys employing many filters produce high dimensional data sets which cannot be so easily visualised, and probably contain much more information than low dimensional slices through them. Thus the HFD results might be telling us that overlapping filters provide a more efficient sampling of stellar spectra than non-overlapping systems. This is not implausible, also given that each filter probably has a different relevance for determining each AP for each star, so the effective number of filters in each case is reduced.

When compared with alternative filter systems proposed for the same instrument, the HFD system performs much better in terms of overall fitness. This is largely due to the broad filters and hence larger AP-gradients than the conventional systems. One might think therefore that the HFD system suffers in terms of the vector separation, as broad filters may tend to 'wash out' the signature of individual APs. But an analysis of the distribution of the vector separation terms shows that this is not the case (Bailer-Jones 2004). Nonetheless, it is found that both the HFD and conventional systems continue to suffer from some serious limitations, in particular relating to the vector separation for particular APs. This will be addressed in future developments of the model.

It is interesting to follow the evolution of the filter system parameters during the optimization, as shown in Fig. 5. Looking first at the central wavelength (top left) we see that the filters occupy a fairly broad part of the parameter space for the first 20 or so iterations (generations). (Filters at the longest or shortest wavelengths are less common, but this is because very broad filters – which are present early on – cannot have short or long central wavelengths due to the wavelength bounds.) After about 20 iterations, some clear preferred regions appear which continue throughout the optimization and the region between 6000 and 8000 Å is disfavoured throughout. Turning to the filter width (top right) we see that, although filters with a HWHM up to 4000 Å are permitted, after about 20 iterations the population is largely purged of filters wider than 2000 Å. A few dominant regions narrower than this stand out, but generally a range of widths are represented. The evolution of the fractional integration time (bottom panel) is quite different. They are initialized to equal values yet quickly diverge to cover the full range possible. Note the gap at low t. This is because a lower limit of 0.025 was imposed for practical reasons: filters allocated very little time will be ineffective due to low SNR. If a mutation takes t below 0.025 then t is set to zero (the thick

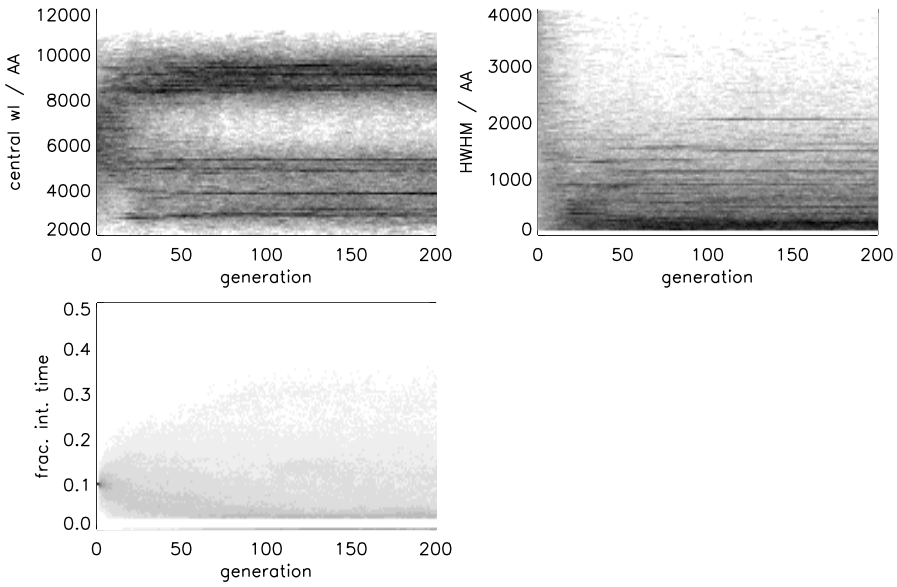

Fig. 5. Evolution of the three filter system parameters for a 10-filter system. At each generation there are ten points (one per filter) for each of the 200 filter systems for that filter parameter type, plotted as a grey scale. The corresponding fitness evolution is shown in Fig. 3 and the best resulting system is that shown in Fig. 4.

line at the bottom). A positive mutation turns a filter on again. A maximum value of t of 0.4 is also imposed yet we see that HFD essentially self imposes a more stringent limit of about 0.3. Clearly it is inefficient if any one filter severely dominates the integration time budget.

4 Conclusions and future work

The Heuristic Filter Design model represents a systematic way for designing photometric filters by casting this as a formal optimization problem. This makes it amenable to the extensive optimization literature. The current model is somewhat rudimentary, yet produces filter systems which are competitive with other systems designed for the same problem/instrument, at least according to the figure-of-merit developed here. The filters are somewhat unconventional – broad and overlapping – yet physically we can see why this may be preferred. Nonetheless, a number of improvements should be made to the model. First, the fitness function may be an oversimplification: it only accounts for linear variations in the data space and ignores any global degeneracies. It is also prone to 'overseparate' some stars or APs at the expense of others. Part of the problem here is that the fitness is a combination

of fundamentally different terms with different scales, so the optimization is dependent on the weighting adopted (not discussed here; see Bailer-Jones 2004). One way around this might be to use multiobjective optimization methods. In addition, more sophisticated genetic operators for search and selection could be employed, e.g. to make the search more directed, perhaps by explicitly incorporating astrophysical information.

References

BAILER-JONES, C.A.L. (2002): Automated stellar classification for large surveys: a review of methods and results. In: R. Gupta, H.P. Singh, C.A.L. Bailer-Jones (Eds.): *Automated Data Analysis in Astronomy*. Narosa Publishing House, New Delhi, 83–98.

BAILER-JONES, C.A.L. (2004): Evolutionary design of photometric systems and its application to Gaia. *Astronomy & Astrophysics*, in press

BÄCK, T. and SCHWEFEL, H.-P., 1993: An overview of evolutionary algorithms for parameter optimization. *Evolutionary Computation*, 1, 1–23