Heuristic design of filter systems using an evolutionary algorithm

Coryn A.L. Bailer-Jones

Max-Planck-Institut für Astronomie, Heidelberg

Carnegie Mellon University, Pittsburgh

PWG Meeting, Leiden, 9-10 October 2003

thanks to Jos de Bruijne for helpful discussions and comparisons

Filter design problem

... to design a filter system to

simultaneously determine

multiple astrophysical parameters (APs)

across a *wide* parameter space

subject to instrumental *contraints* ...

Considerations

- multiple conflicting demands on filter system
- modifications of existing filter systems (designed for more restricted goals) is not necessarily appropriate
- manual design complex, and gives no idea of optimality (is there a better system subject to same constraints?)
- cast as a mathematical optimization problem:
  - parametrize filter system
  - establish a figure-of-merit of filter system performance
  - maximise this as a function of the filter system parameters
  - pure gradient methods get stuck in local minima

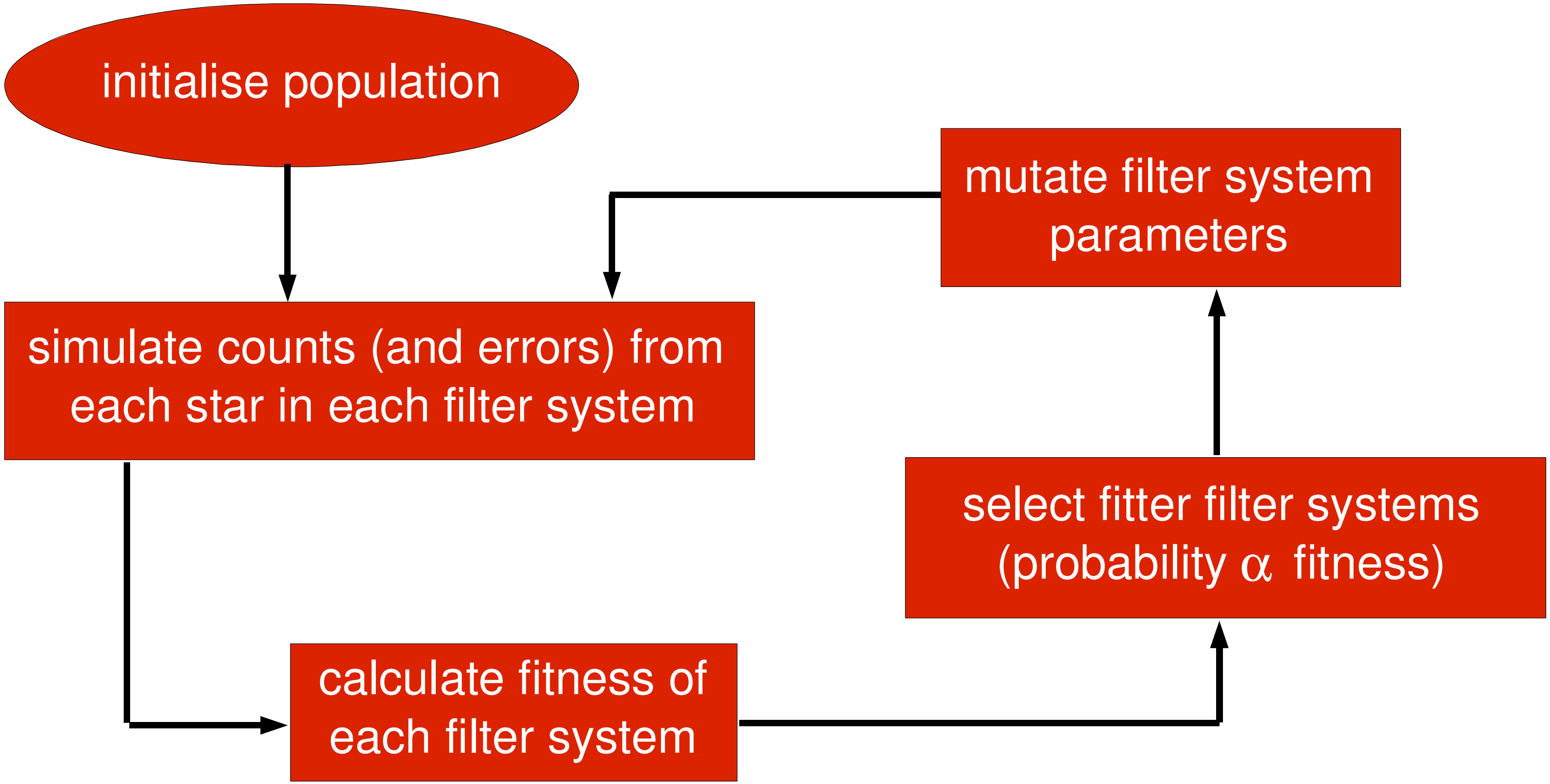
Evolutionary algorithms

- population-based methods overcome local optima and permit a more efficient search of the parameter space
- 1 individual in population = 1 candidate filter system
- Evolutionary Algorthms (EAs) use the principle of natural selection from biological evolution
  - Genetic Algorithms (GAs), Evolutionary Strategies (ESs), Evolutionary Programming (EPs)
- genetic operators
  - reproduction: recombination; mutation (exploration)
  - selection (exploitation)
- provides a stochastic (but not random) search
- population evolves towards optimum (or optima)

Heuristic filter design (HFD) model

- figure-of-merit / fitness function:
  - solving a full regression model to test every filter system too slow and unnecessary
  - construct a measure of ability of filter system to maximally “separate” stars with a range of APs represented by a (synthetic) grid
  - grid showing variance in T_{eff}, logg, [Fe/H], A_V
  - use a simple instrument model to simulate counts and errors in candidate filter systems
- fixed instrument parameters and number of filters (but number of “effective” filters can vary)
- evolve population and find fittest filter system

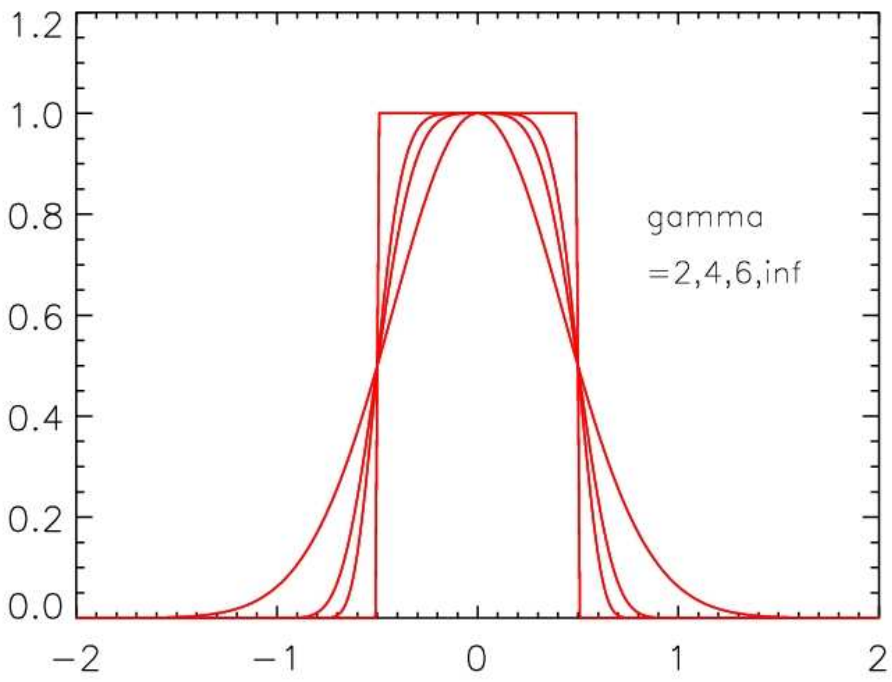
HFD model


initialise population
mutate filter system parameters
simulate counts (and errors) from each star in each filter system
select fitter filter systems (probability α fitness)
calculate fitness of each filter system

Filter system representation

Each filter system consists of *I* filters each with 3 parameters:

c central wavelength
w half width at half maximum
t fractional integration time (of total available for all filters)

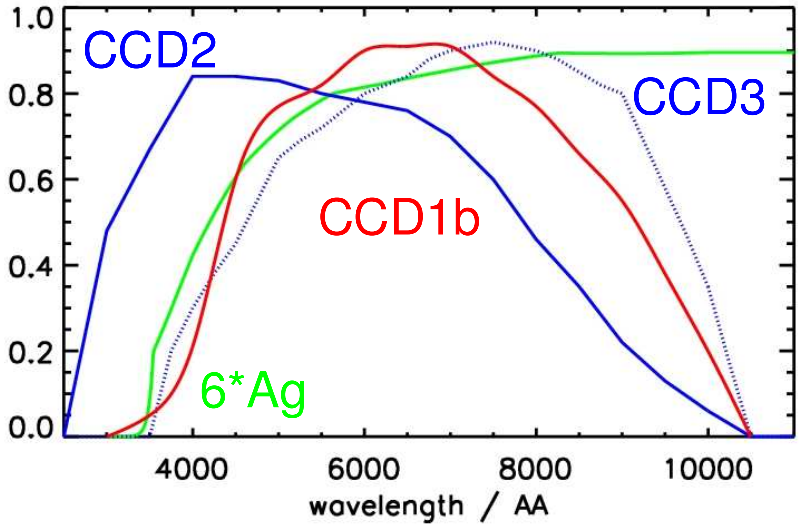


Generalised Gaussian profile with $\gamma = 4$

$$y = \exp(-\ln 2\ [(\lambda - c)/w]^{\gamma})$$

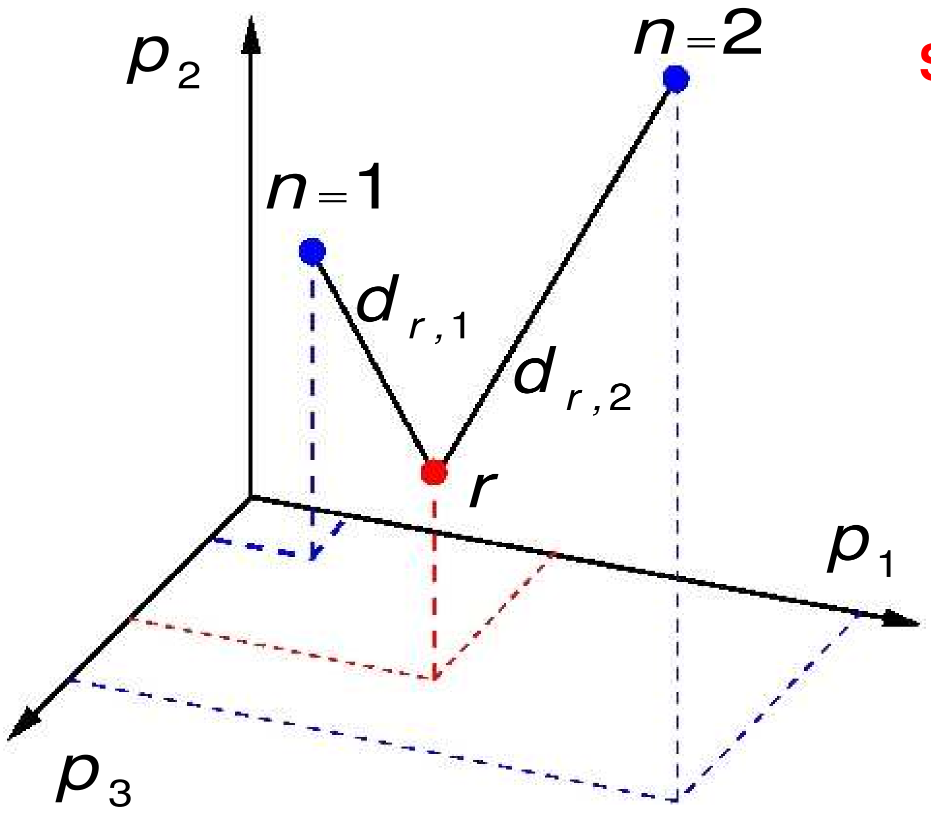
Instrument model

- photon count model (photsim)
- end 2002 instruments (CUO-116)
- simple aperture photometry
- source, sky and R/O noise
- end-of-mission mean values



		BBP	MBP
Telescope aperture area	m^2	0.7	0.25
Total integration time	s	1200	16500
Instrument response		6*Ag	3*Al
CCD response		CCD1b+3	CCD2+3
CCD readout noise	e-	205	226
Effective background	G mag	23.12	19.05

Fitness: SNR distance



SNR distance of star r from neighbour n:

$$d^2_{r,n} = \sum_i \frac{(p_{i,n} - p_{i,r})^2}{\sigma^2_{i,n} + \sigma^2_{i,r}}$$

$p_{i,n}$ = photon counts in filter *i* for star *n*

$\sigma_{i,n}$ = expectation of error in $p_{i,n}$

photon counts (and errors) are area normalised, i.e. $\Sigma_i \, p_{i,n} = 1$
(generalization of forming colours)

Fitness: AP gradient

SNR-distance does not take account of the APs.
But the **AP gradient** does:

$$\frac{d_{r,n}}{\Delta\phi_{j,n,r}}$$

$\Delta\phi_{j,n,r}$ = difference in AP j between star n and r

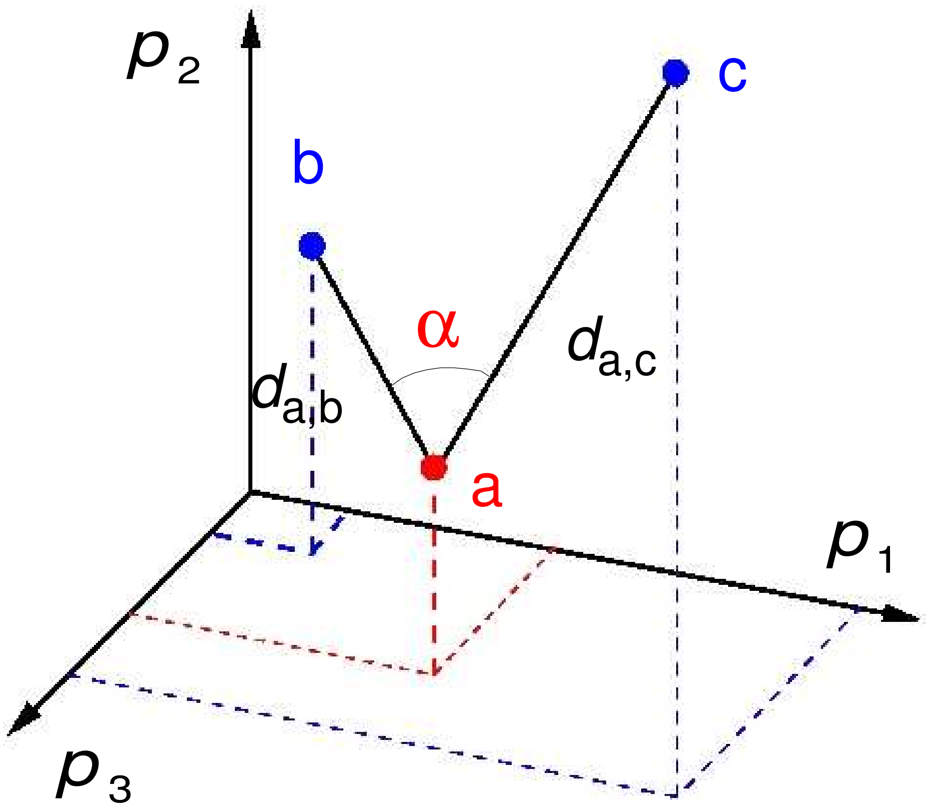
This is appropriate for a single AP but obvious generalization to multiple APs does not work:

$$\sum_j \frac{d_{r,n}}{\Delta\phi_{j,n,r}}$$

Extension to multiple APs is more complex. Must allow for:

- simultaneous effects of APs on data
- degenerate effects of APs on data
- very different magnitude effects on data (eg. T_{eff} vs. [Fe/H])

Fitness: vector separation



For each source, a, and each AP, *j*, find nearest neighbour (NN) which differ only in *j* ("isovars"), e.g. b and c

Calculate angle, α, between vectors: Nearer to 90° => better separation (less degeneracy)

Calculate magnitude of **cross product**:

$$V_{a,b,c} = d_{a,b}\, d_{a,c} \sin\alpha$$

N.B. vectors are SNR-weighted, i.e. if $\mathbf{c} = (c_1, c_2, \ldots c_i, \ldots)$ then

$$c_i = \frac{p_{i,c} - p_{i,a}}{\sqrt{\sigma_{i,c}^2 + \sigma_{i,a}^2}}$$

Fitness: final measure

Cross product: $V_{a,b,c} = d_{a,b}\, d_{a,c} \sin\alpha$

Now use concept of AP-gradients
(plus weighting of APs to boost significance of weak APs esp. [Fe/H] and logg)

$$f_{a,b,c} = V_{a,b,c} \frac{w_b \, w_c}{\Delta\phi_{a,b} \, \Delta\phi_{a,c}}$$

weight for AP j, i.e. that which differs between a and c

difference in AP j between a and c

J APs => J NNs per source => J(J-1)/2 cross products per source

$$\boxed{\text{Fitness} = \sum_a \sum_{b,c} f_{a,b,c}}$$

sum over all NNs for source
and for all sources in grid

Genetic operators

Selection

Individuals from parent population (generation g) selected (with replacement) with probability proportional to fitness. *Elitism* used to guarantee selection of best few.

=> *intermediate population (IP)*

Mutation

Parameters of each individual $\{c_i, w_i, t_i\}$ in IP mutated with a finite probability:

$c_i\,(g+1) = c_i\,(g) + N(0,\sigma_c)$

$h_i\,(g+1) = h_i\,(g)[1 + N(0,\sigma_h)]$ => children (next generation)

$t_i\,(g+1) = t_i\,(g)[1 + N(0,\sigma_t)]$

Stellar grid

log g	T_{eff} / K (SpT)						
4.5	3500 MV	4750 KV	5750 GV				
4.0				6750 FV	8500 AV	15000 BV	35000 OV
3.5							
3.0				6000 RRLyr	8500 BHB		
2.5			5500 GIII			15000 BIa	
2.0		4500 KIII	5500 FIa				
1.5					8500 AI		
1.0	3500 MIII		5000 GI				
0.5							
0.0	3500 MIa						
	[Fe/H]:		+0.5	0.0	−0.5	−1.5	−2.5
	A_V:		0.0	0.2	2.0	5.0	10.0

Purpose:
to represent
how data depend
upon APs

17 T_{eff} / logg combinations at each of 5 [Fe/H] and extinction values
=> 425 sources

BaSeL 2.2 library
+ Fitzpatrick (1999) extinction curves

noise-free data

HFD application

Free parameters: 3*I

central wavelength, c / Å
half width at half max., h / Å
fractional integration time, t

EA parameters

size of population, K	200
size of elite	10
number of generations	200
probability of mutation	0.4
σ[c] min./max. / Å	25/500
σ[h] min./max.	0.05/0.5
σ[t] min./max.	0.05/0.5

Strategy parameters: 3*K

σ[c], σ[h], σ[t] per filter system

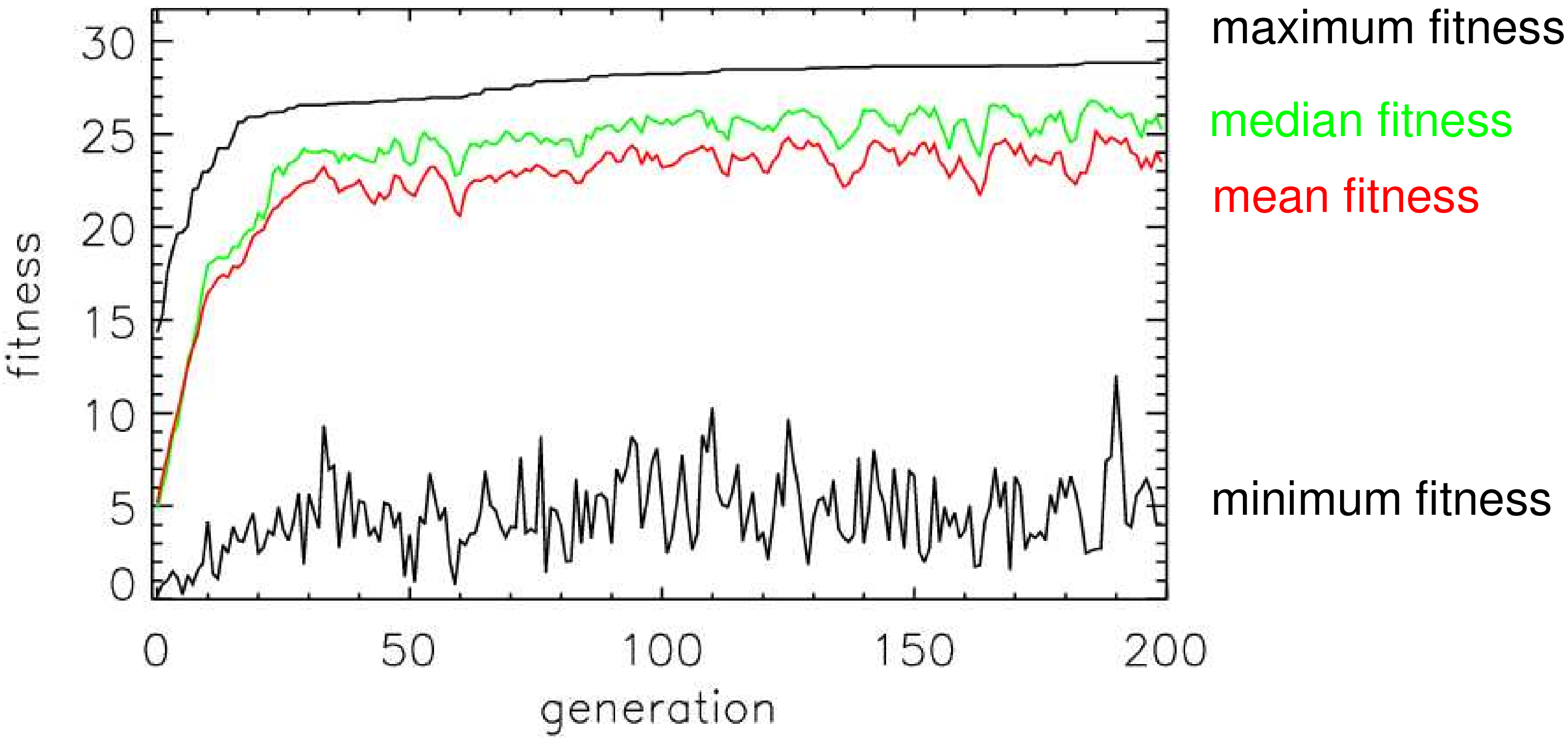
Limits of search domain:

- λ limits from CCD/instrument QE
- 80 Å < h < 4000 Å
- no limits on t (just normalization)

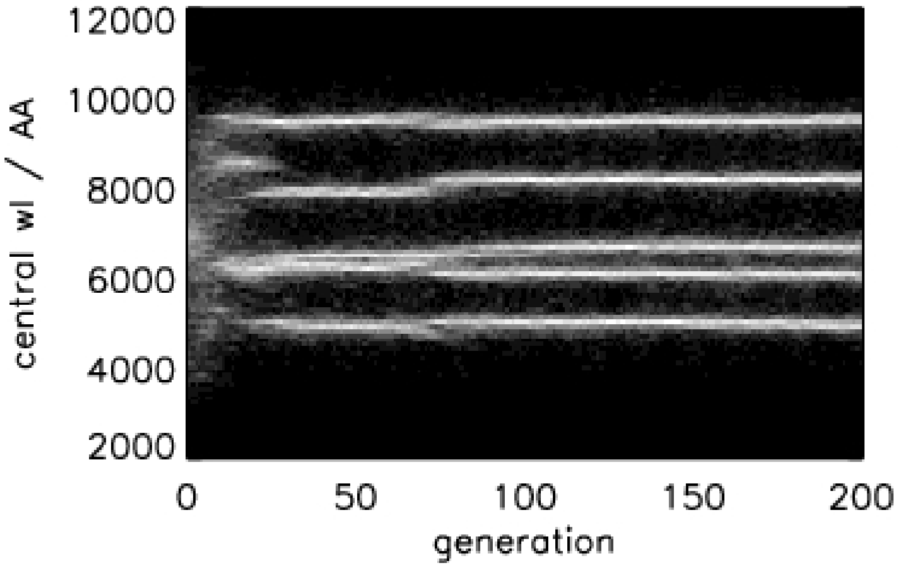
Applications:

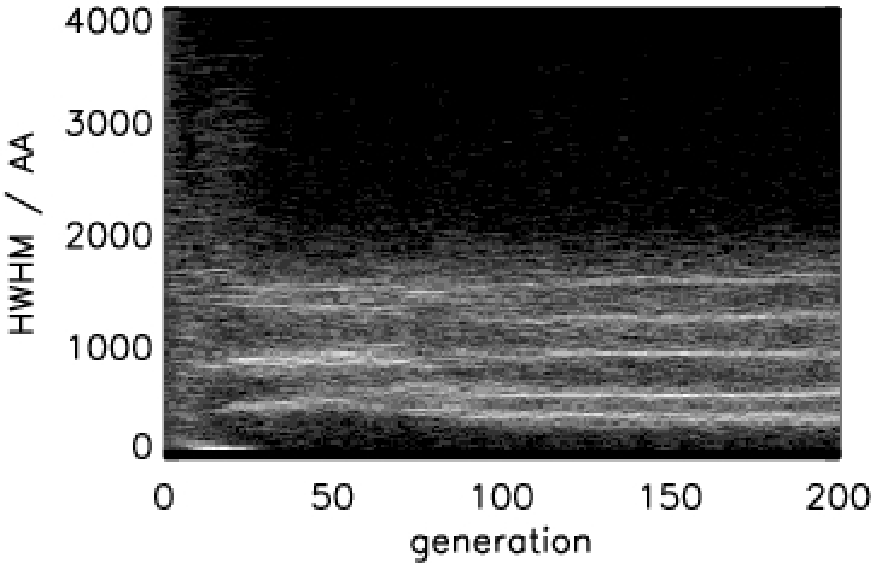
1. BBP with 5 filters (G=20)
2. MBP with 10 filters (G=20)

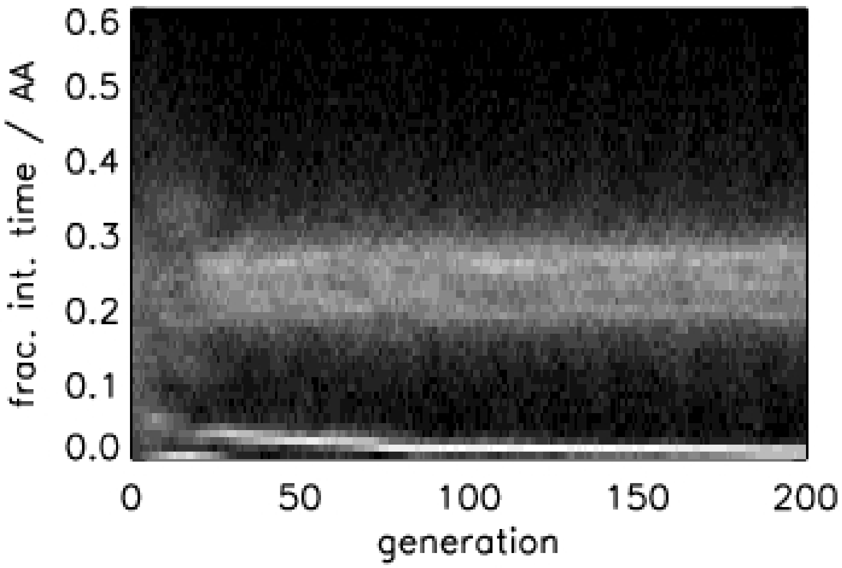
BBP-5 system


maximum fitness
median fitness
mean fitness
minimum fitness
fitness
generation
0
5
10
15
20
25
30
0
50
100
150
200

BBP-5 system

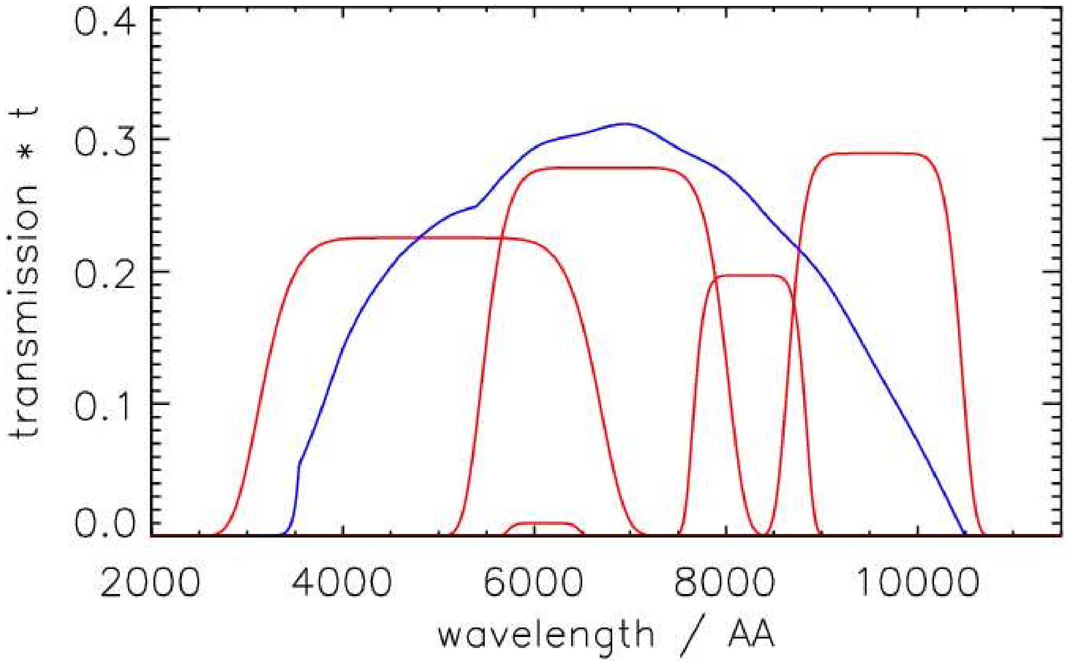






Evolution of all filter system parameters
(200*5 for each parameter type at each generation)

BBP-5 system

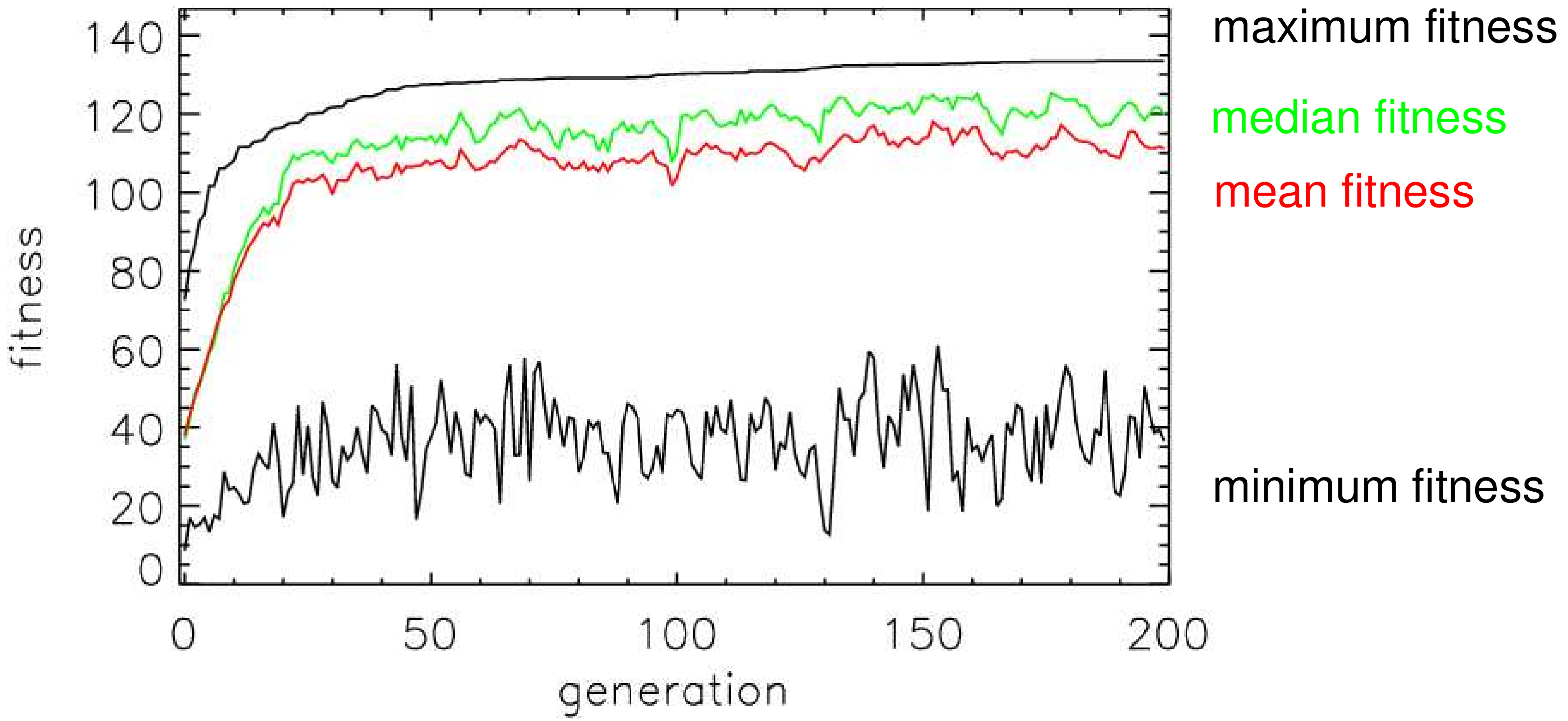


red = filter transmission x fractional integration time

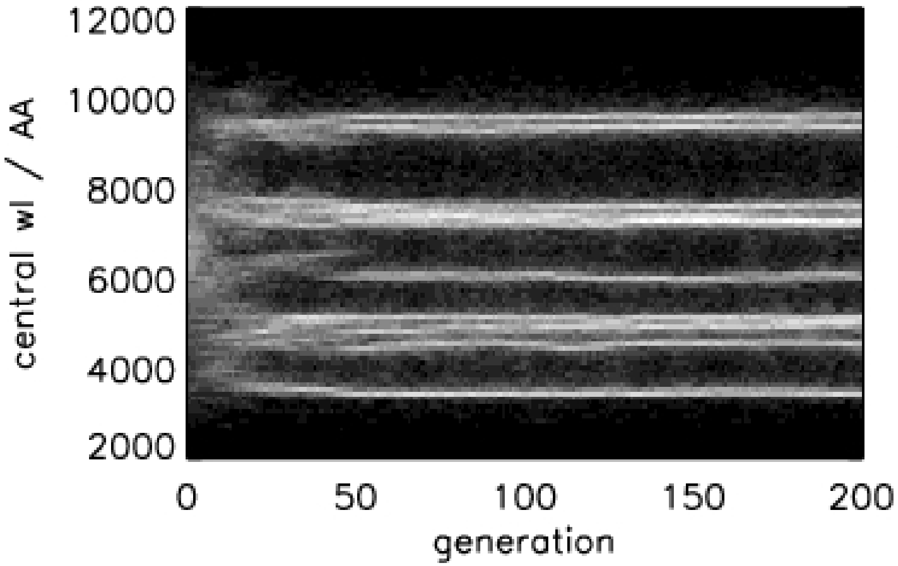
blue = Instrument*CCD QE (scaled)

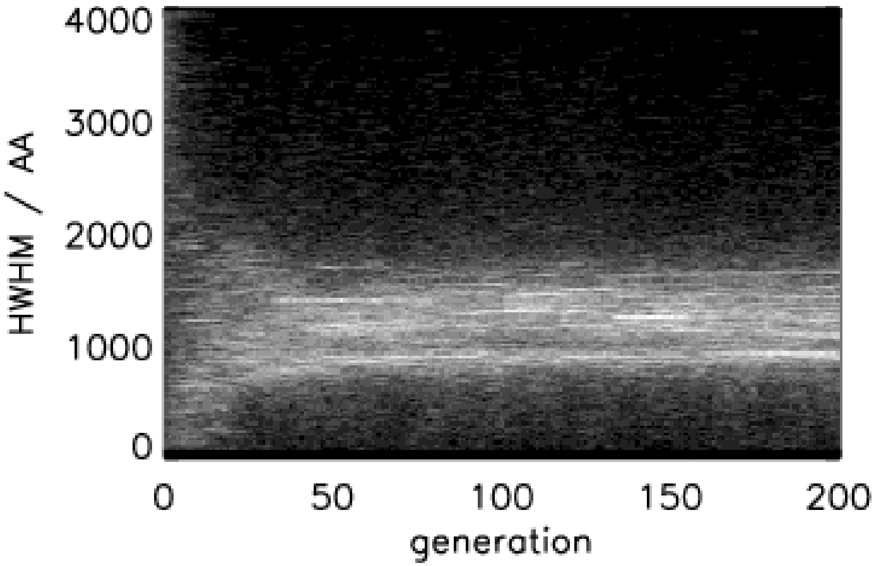
- broad filters (bright limit: fitness α sum photons collected)
- overlapping filters
- 4 effective filters (1 almost “turned off”)

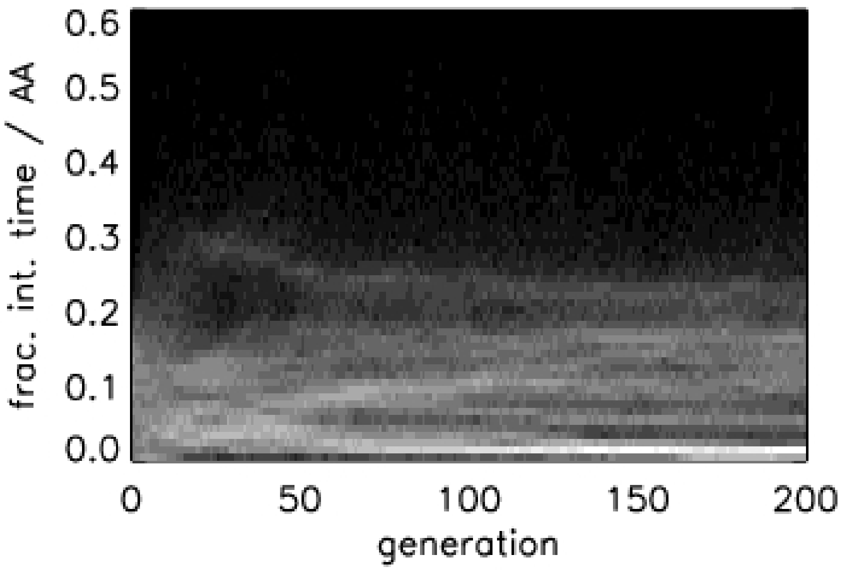
MBP-10 system



MBP-10 system

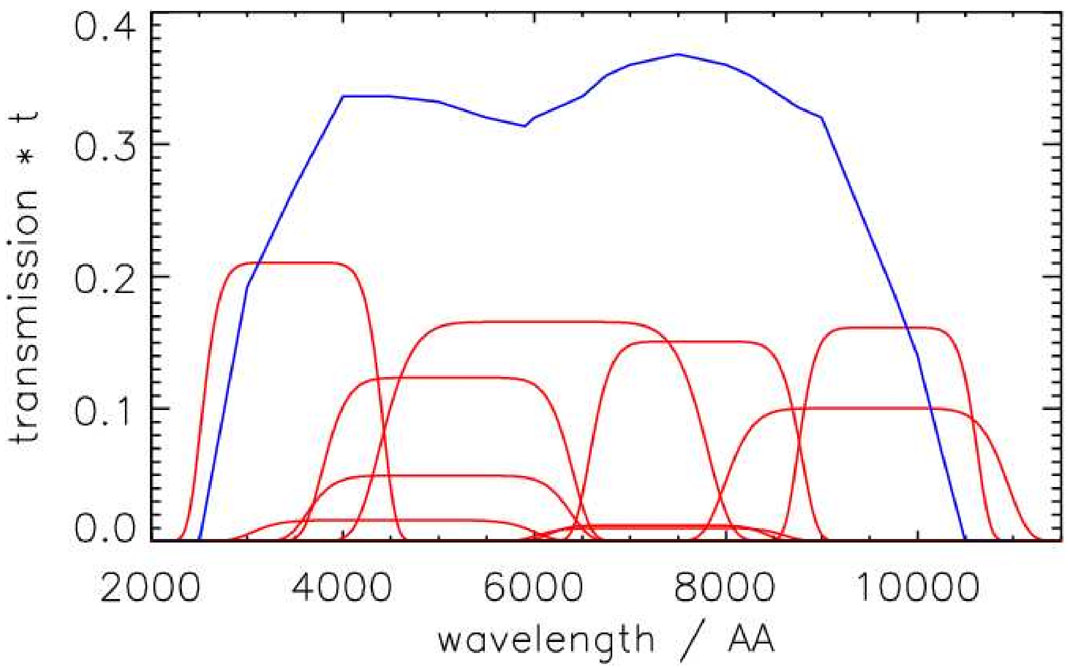






Evolution of all filter system parameters
(200*10 for each parameter type at each generation)

MBP-10 system

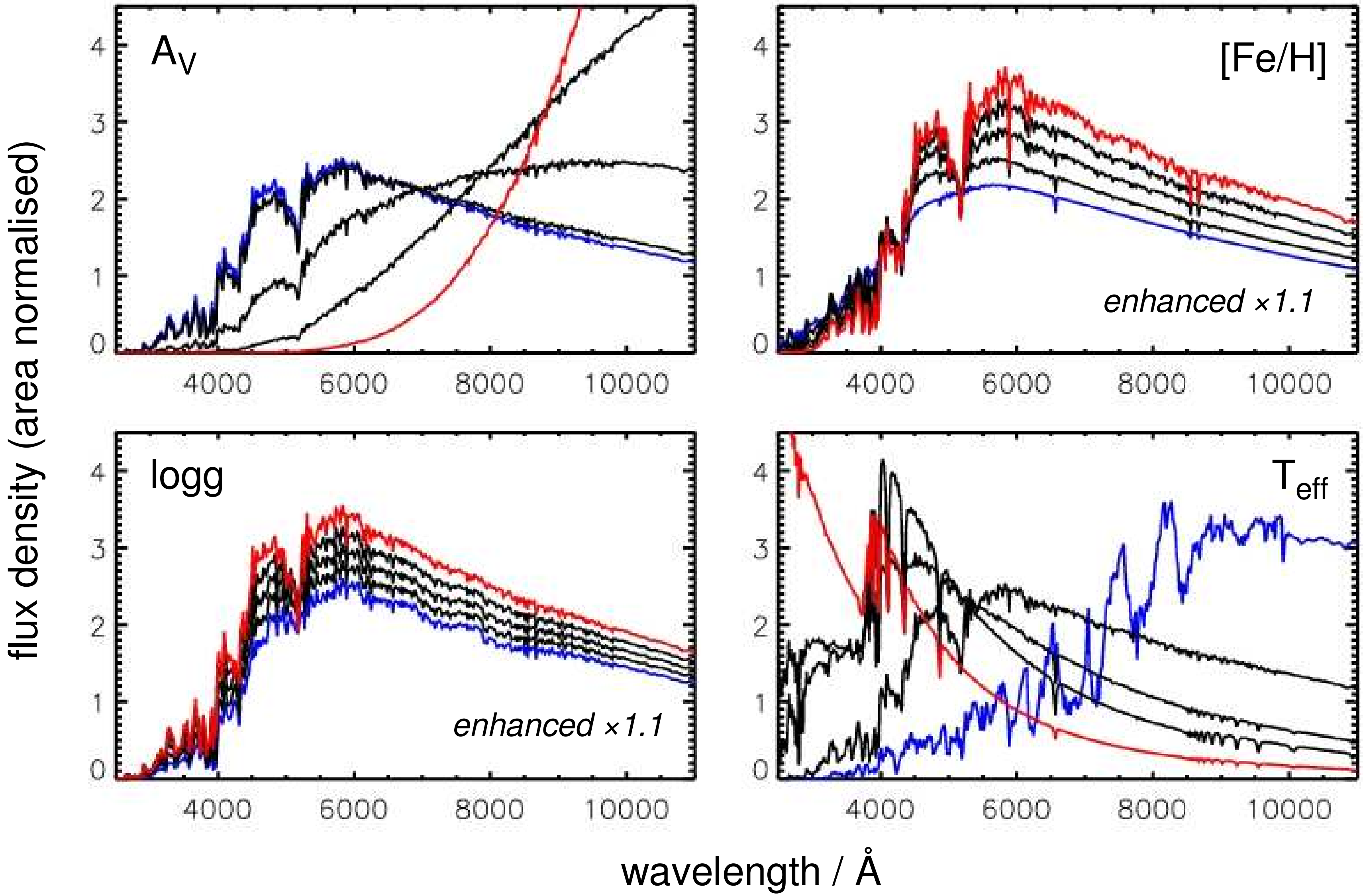


red = filter transmission
x fractional
integration time

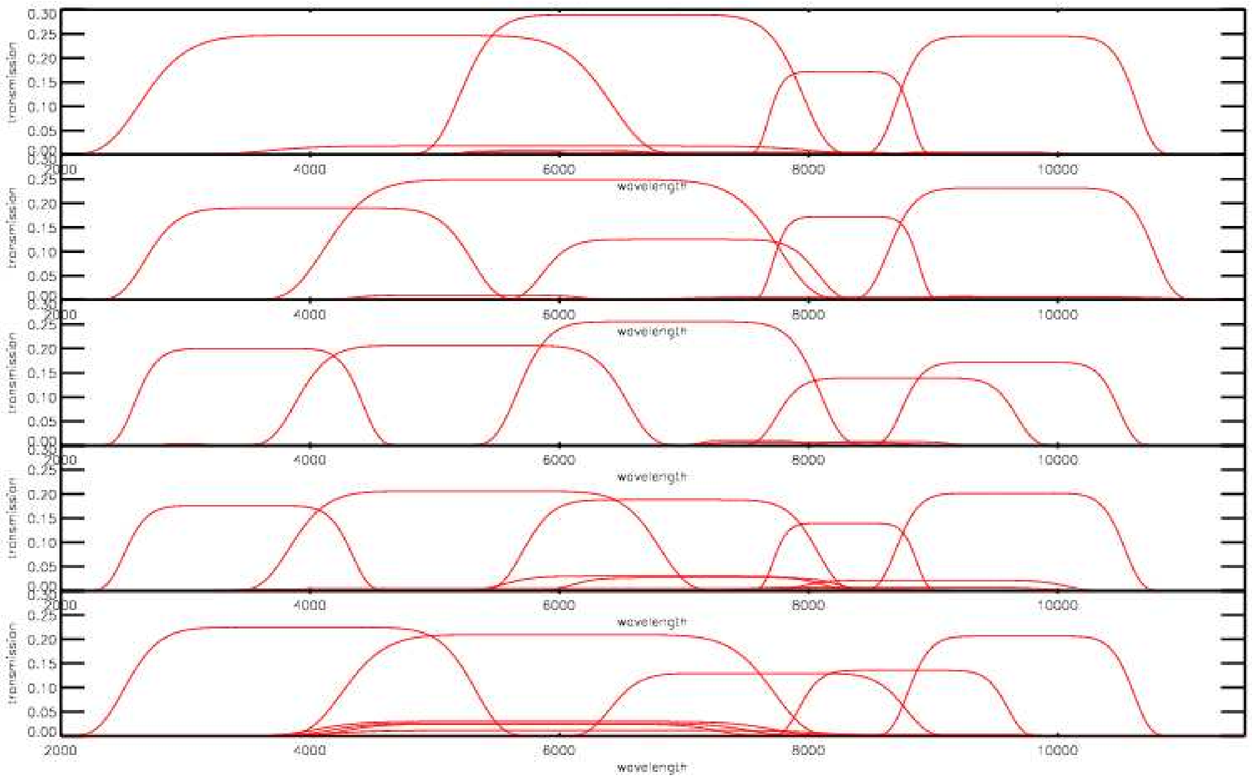
blue = CCD QE (scaled)

- broad, overlapping filters
- 6-8 effective filters:
  - 2 almost identical to 2 others
  - 3 almost “turned off” (t = 0.01-0.015, but broad, so may not be irrelevant)
  - Cf. 4 APs requiring minimum of 5 filters

AP variations have broad band effects



HFD reproducibility



Best MBP-10 filter system found in each of 5 runs with different initial filter systems

- MBP: often 5 effective filters
- BBP: two set of solutions with 3 and 4 effective filters
- good fitness repodocibility in best FSs (<10% variation for BBP and MBP)
- reasonable FS reproducibility (esp. with few filters or APs, small grid)

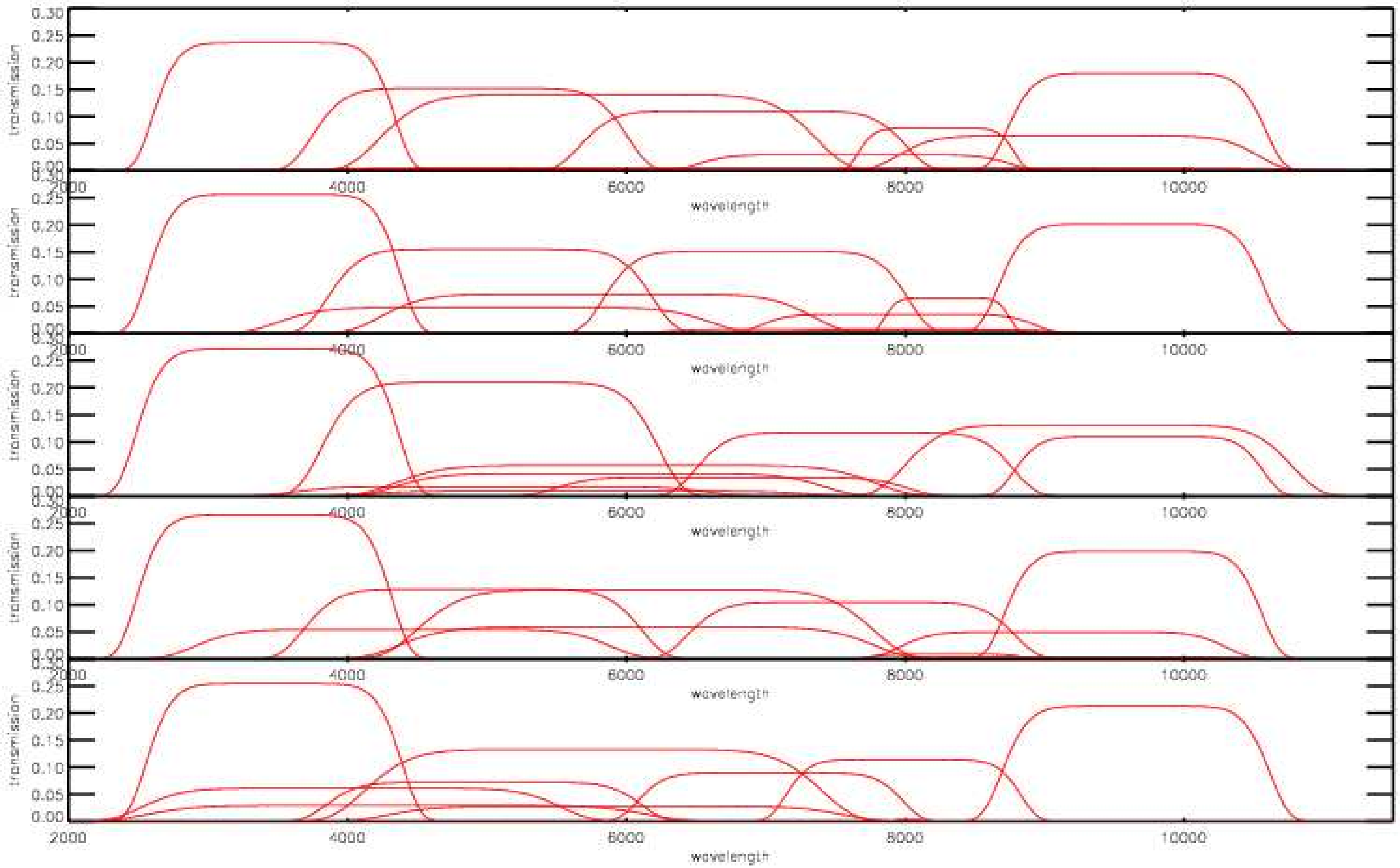
Effects of modifications

Change to BBP setup	**Effect**
restrict fitness sum to T_{eff} and A_V	3 broad, overlapping filters covering whole wavelength range

Change to MBP setup	**Effect**
restrict fitness sum to T_{eff} and A_V (grid remains unchanged)	more time allocated to bluest filter
restrict grid to $T_{eff} < 8000$ K	much broader blue filter
all sources at G=15	many more (~ 8) effective filters*

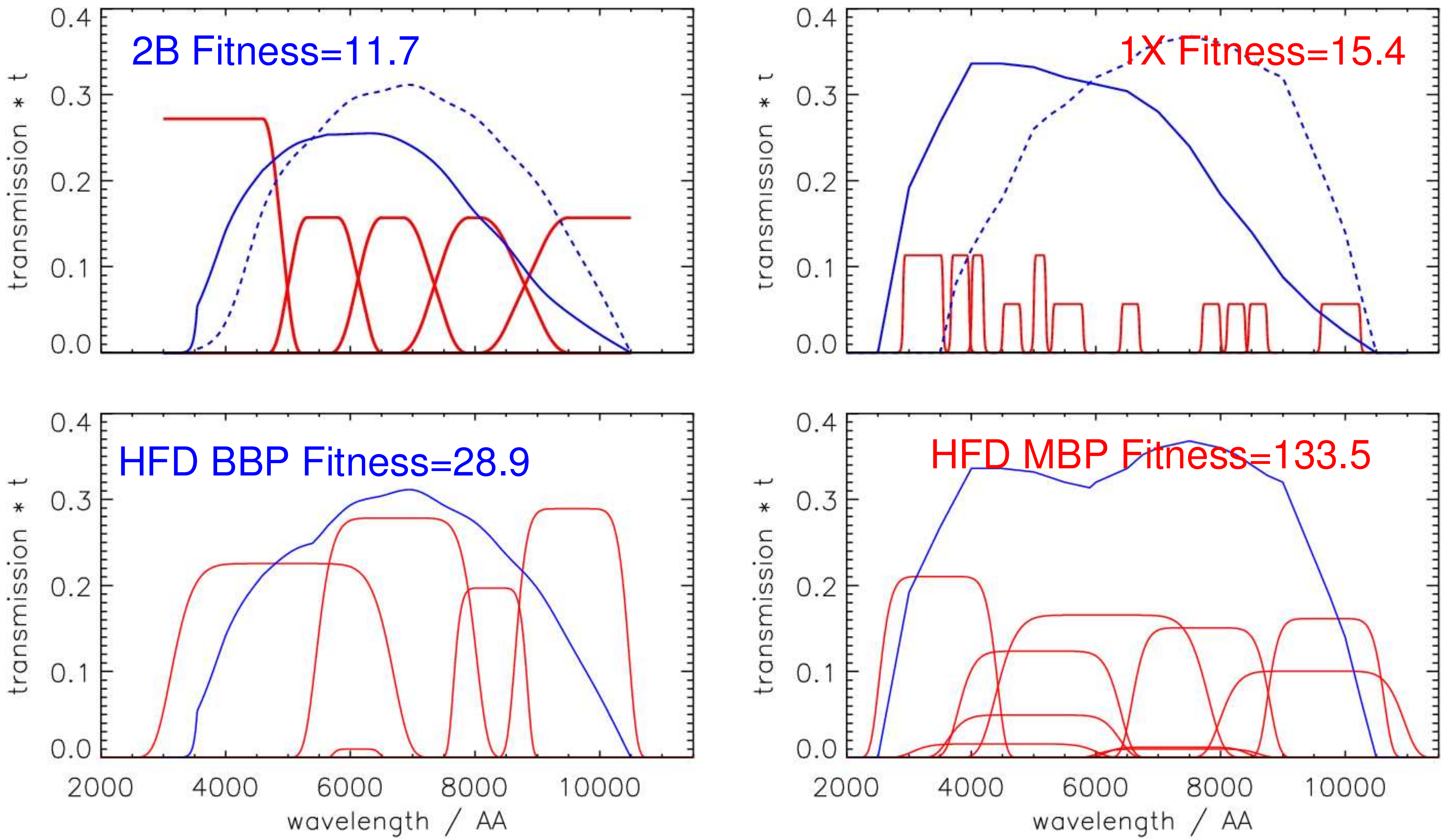
*Faint limit: fitness α $1/\sqrt{N}$ (N = no. filters), relevant for MBP G=20
Bright limit: fitness independent of N, relevant for MBP G=15

MBP-10 at G=15



transmission
wavelength
0.30
0.25
0.20
0.15
0.10
0.05
0.00
2000
4000
6000
8000
10000

Comparison with 2B and 1X



Conclusions

- Filter design principles
  - vector separation figure-of-merit to maximise SNR-distance between all sources and minimise degeneracy between APs
  - evolutionary operators to efficiently search filter parameter space
- HFD gives a strong preference for:
  - broad filters (AP signatures spread over large part of spectrum)
  - overlapping filters covering whole wavelength range

  => better use of higher dimensional data space and limited resources
- much higher fitnesses achievable than 2B, 1X or 2F (with this fitness function and grid)
- HFD filters should permit simple AP-estimation models

Future work

- Short term
  - get feedback, esp. on fitness function
  - some effects need to be investigated:
    - tendency to "remove" filters (fitness dependence at faint magnitudes?)
    - "over-separating" some sources at the expense of others?
  - improve grid
  - verification of filter systems with AP-estimation models (ANN, MDM)
  - proposal of specific MBP and BBP systems
- Long term
  - inclusion of other APs (extension of grid)
  - inclusion of "contaminants"
  - addition of fixed filters (e.g. RVSM, ground-based J,H,K)
  - use of multiobjective optimization methods with EAs