Object classification and the determination of stellar parameters

Coryn A.L. Bailer-Jones

Max-Planck-Institut für Astronomie, Heidelberg

Philip Willemsen, Torsten Kaempf

Sternwarte der Universität Bonn

Classification objectives

To determine the intrinsic properties of individual objects observed by Gaia

- discrete source classification (DSC)
  - star, galaxy, quasar, asteroid, supernova etc.
- (stellar) astrophysical parameter estimation (APE)
  - T_{eff}, log g, [Fe/H], extinction (A_V), [α/Fe], CNO, V_{rot}, activity
- identification of (unresolved) binaries
- identification of new types of objects
- Classification working group (ICAP)

$\Rightarrow$ catalogue of astrometric *and* astrophysical information

Gaia data

Data	Dimensions	No. sources
BBP	5-6 filters	all
MBP	ca. 10 filters	all (minus crowding)
RVS	~350 bins	G<15 (bright for classification quality)
parallaxes		some (sufficient accuracy)
time domain		all

From Gaia data to APs

Coryn Bailer-Jones
Classification with Gaia
Paris, 4-7 October 2004

Data	**Derived quantity**	**Assumption/Model**
normalized SED	Teff, *log* g, [Fe/H], A()	atmospheric model: APs ∈ SED inversion (data) model
G or abs. SED, , A( )	L (Luminosity)	2.5 *log* L – f (G, abs. SED) = A – 5 *log*
L, Teff	R (Radius)	*log* L α 2 *log* R + 4 *log* Teff
log g, R	M (Mass)	*log* g α *log* M – 2 *log* R
M, R, [Fe/H]	t (Age)	evolutionary model

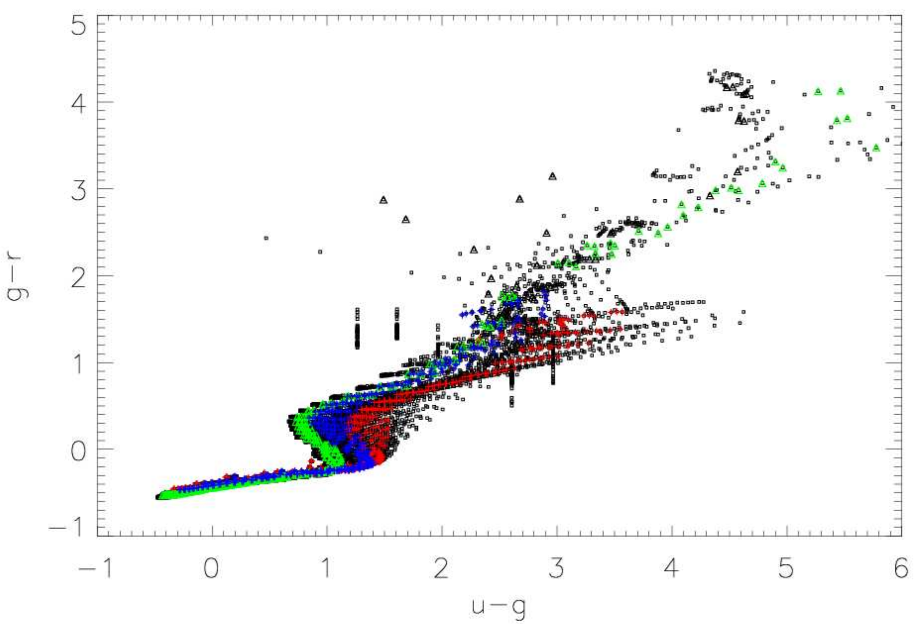
SED = spectral energy distribution, from BBP, MBP, (RVS)
= parallax, A = extinction

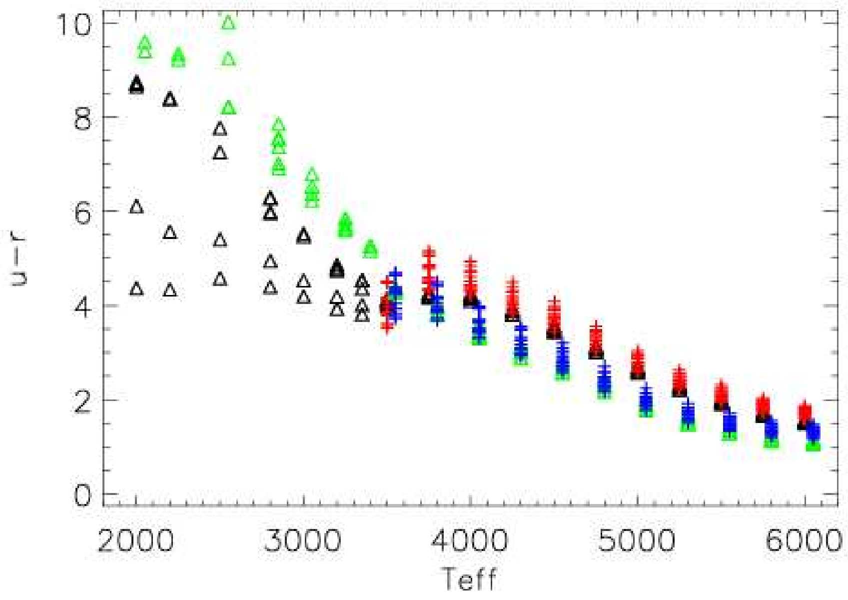
Colour-colour diagrams: SDSS

black triangle = solar [Fe/H] dwarf
green = low [Fe/H] dwarf
red = solar [Fe/H] giant
blue = low [Fe/H] giant
black square = other

Simulated SDSS photometry
of BaSeL 2.2 SEDs

Note: uncalibrated magnitudes
(zero point offset)

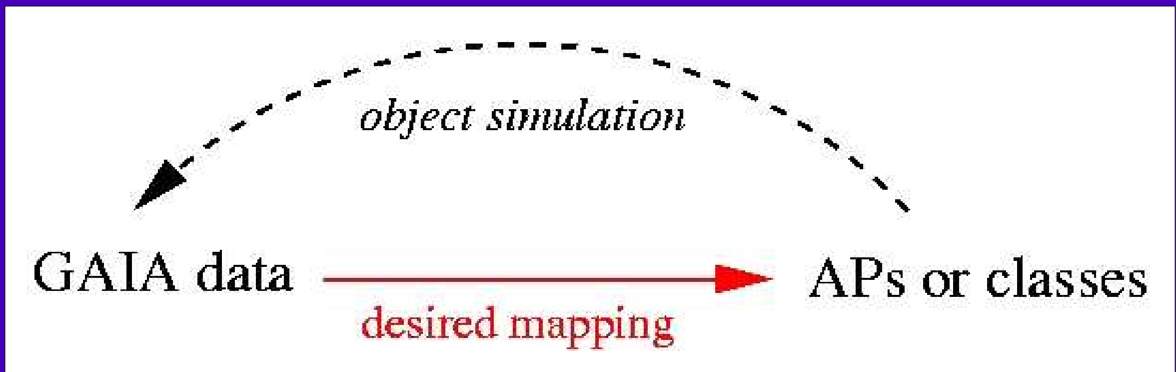




Principle of AP estimation

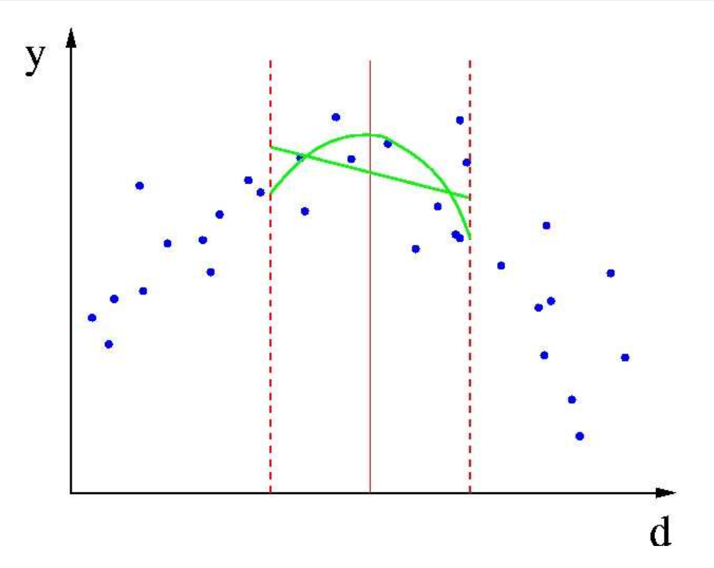
Supervised approach:

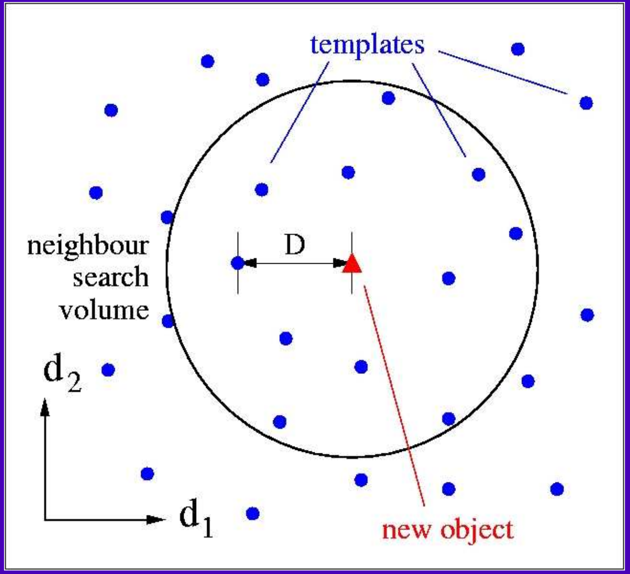
- match data against pre-classified templates (i.e. with known APs)
- synthetic or real template data
- inverse mapping $\Rightarrow$ potential non-uniqueness (AP degeneracy)



Multidimensional regression

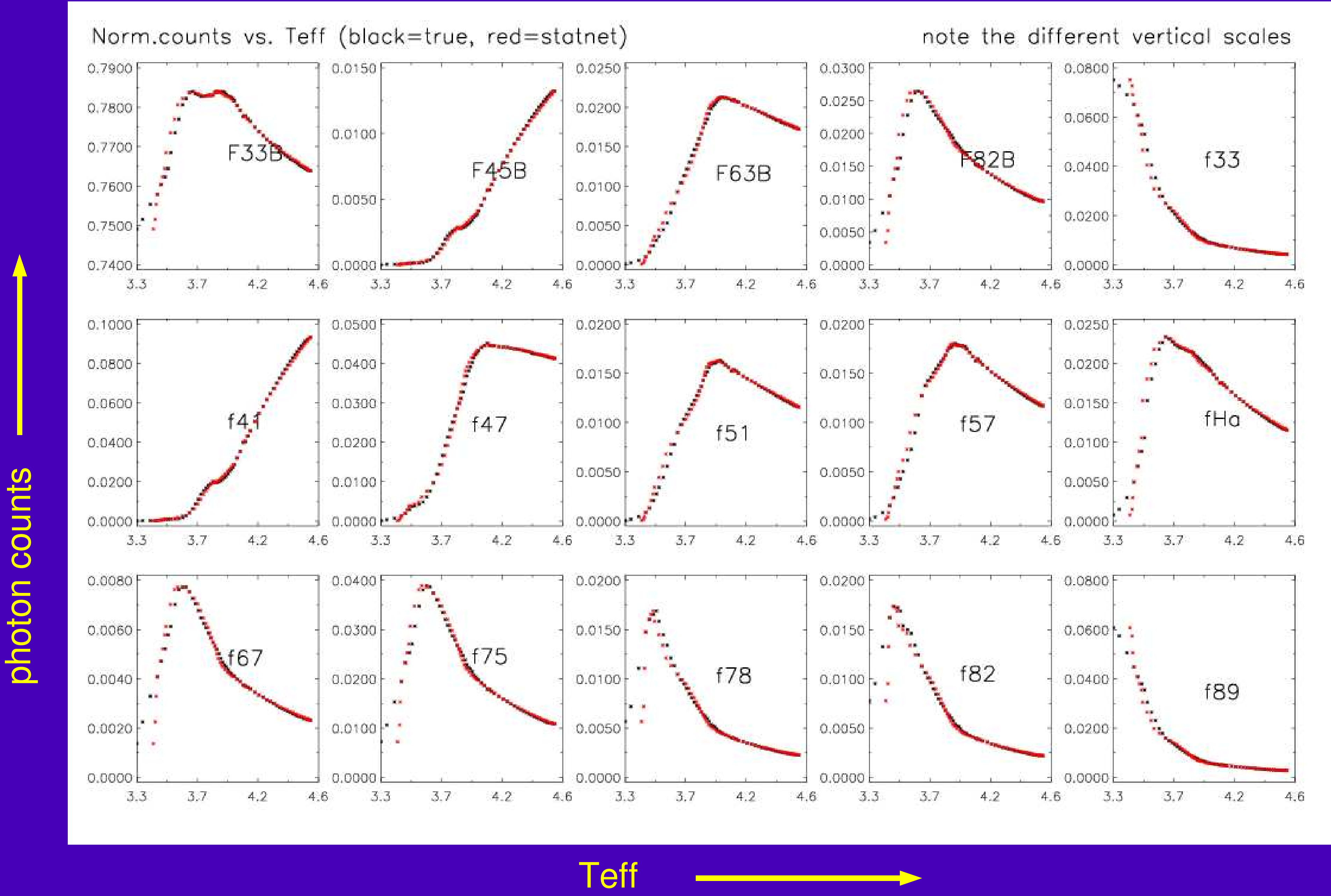
- local vs. global methods
- "curse of dimensionality" ⇔ structured regression
- global, non-linear methods
  - train model on representative data through error minimization
  - e.g. ANN, SVM, MARS
- Examples: ANNs trained/tested on simulated BBP/MBP stellar data to estimate T_{eff}, log g, [Fe/H], extinction (A_V)





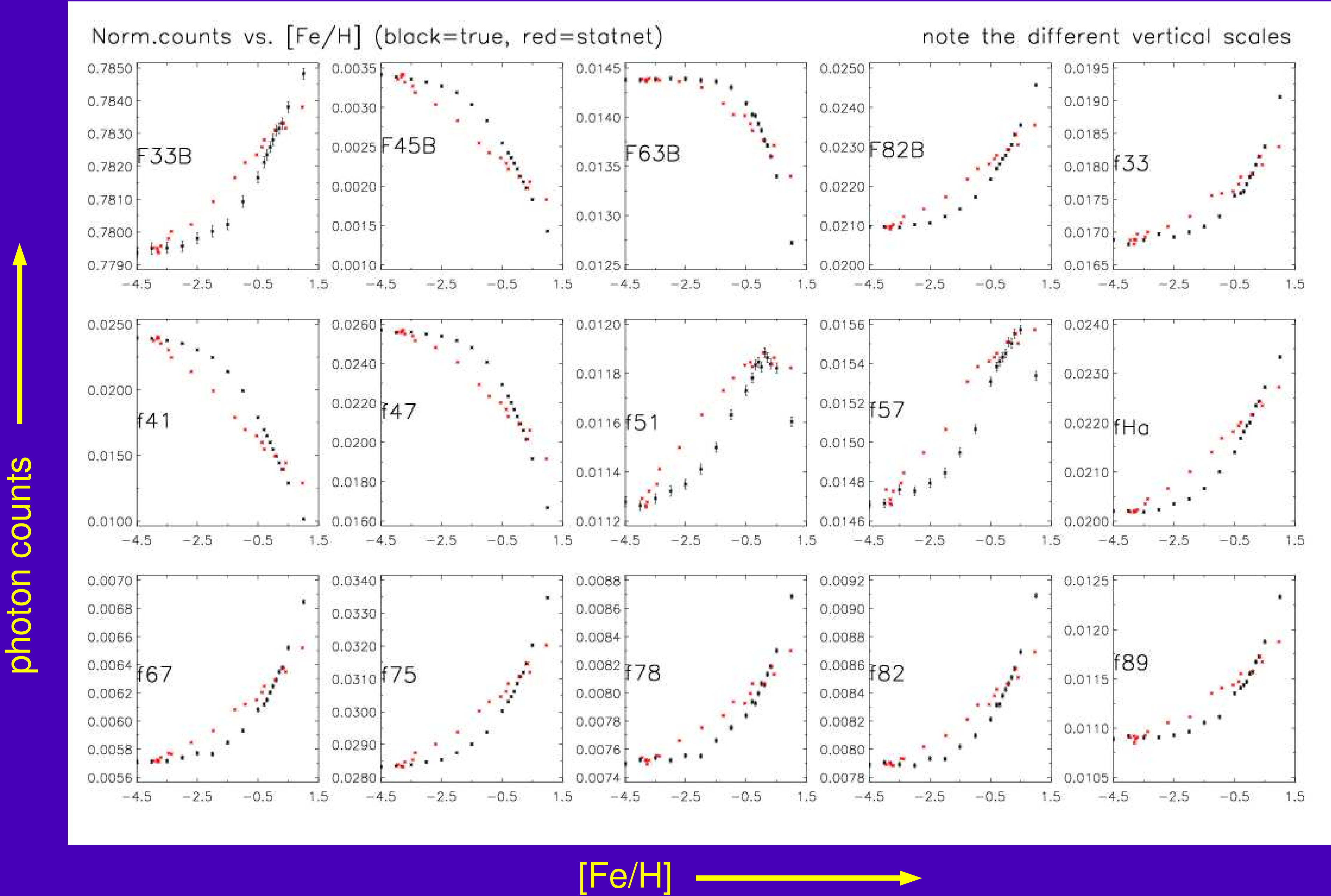
Model performance: T_{eff}

Coryn Bailer-Jones
Classification with Gaia
Paris, 4-7 October 2004



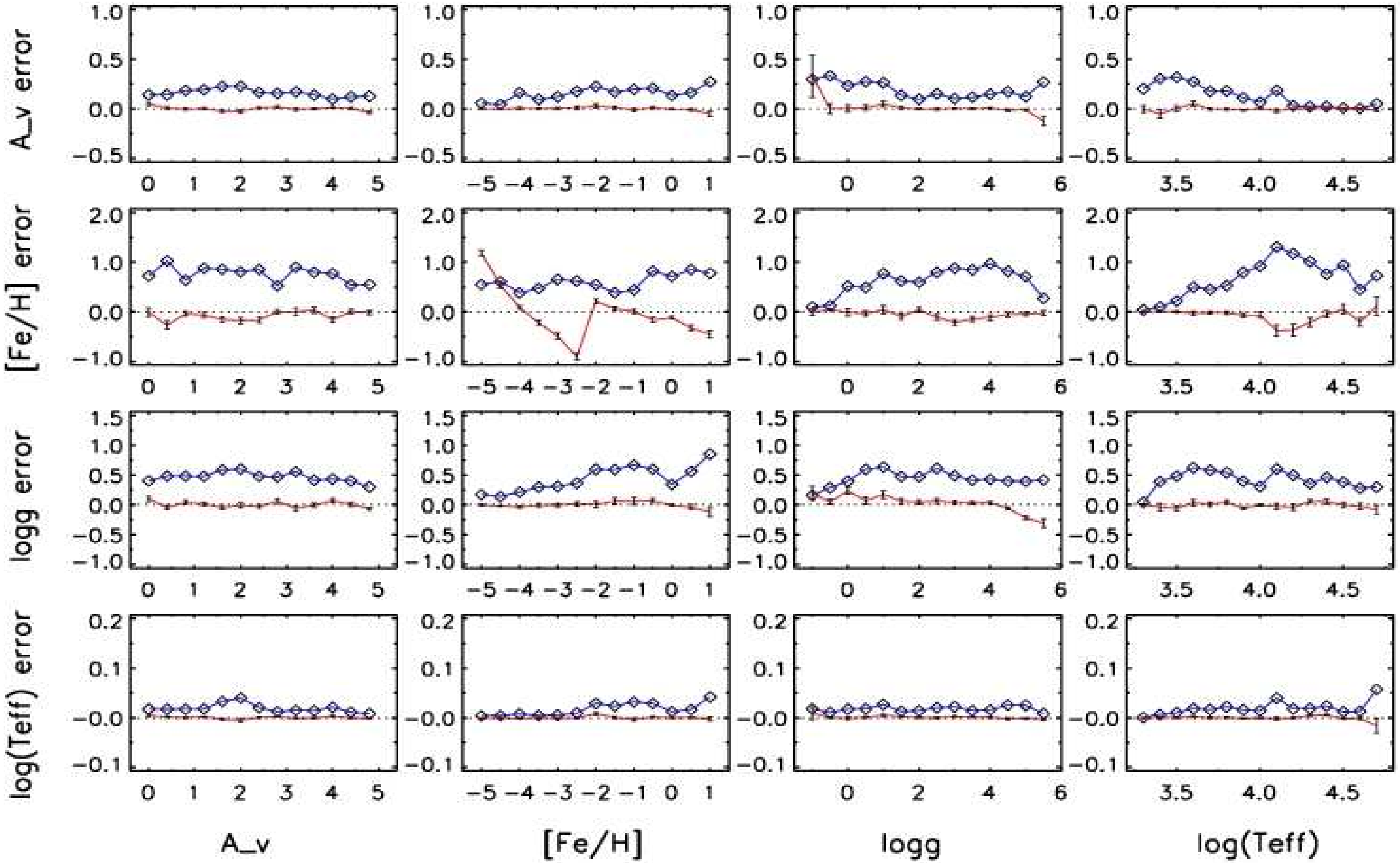
Model performance: [Fe/H]

Coryn Bailer-Jones
Classification with Gaia
Paris, 4-7 October 2004



AP estimation errors

red = mean error (systematic error)   blue = std. dev. about mean (random error)



APE performance from MBP

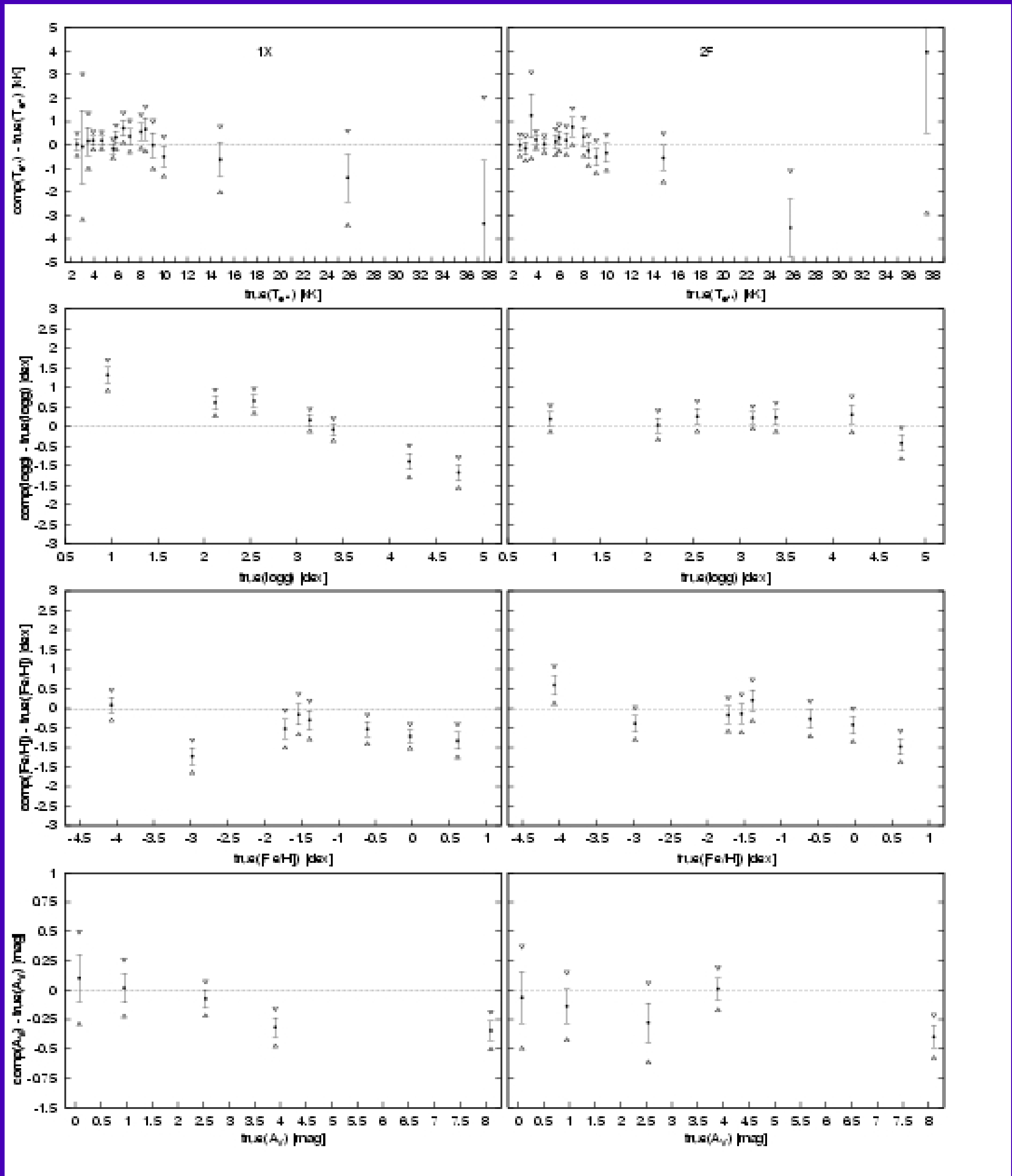
Accuracy varies strongly as a function of the 4 APs and magnitude
⇒ results averaged over wide AP ranges misleading

	$\sigma(A_V)$ mag	$\sigma([Fe/H])$ dex	$\sigma(\log g)$ dex	$\sigma(T_{eff})$ %
KV, G=15, A_V = 0, [Fe/H] = +0.1..-2	~0.1	0.05–0.25	0.2	1–2
KV, G=20, A_V = 0	0.1–0.7	0.1–0.35	0.3–0.5	2–5
KV, G=20, A_V = 6		0.7		8
KIII, G=15, A_V = 0		0.2–0.4	0.2–0.3	2.5–4
KIII, G=15, A_V = 6		0.7–0.8	~1.0	8
AIII, G=15, A_V = 0, [Fe/H] ~ 0	0.08	0.4	0.03	4
BV, G=15, A_V = 6, [Fe/H] ~ 0	0.3		0.35	

Willemsen, Kaempf, Bailer-Jones (2003)

AP error estimates

Coryn Bailer-Jones
Classification with Gaia
Paris, 4-7 October 2004



T_{eff}

log g

[Fe/H]

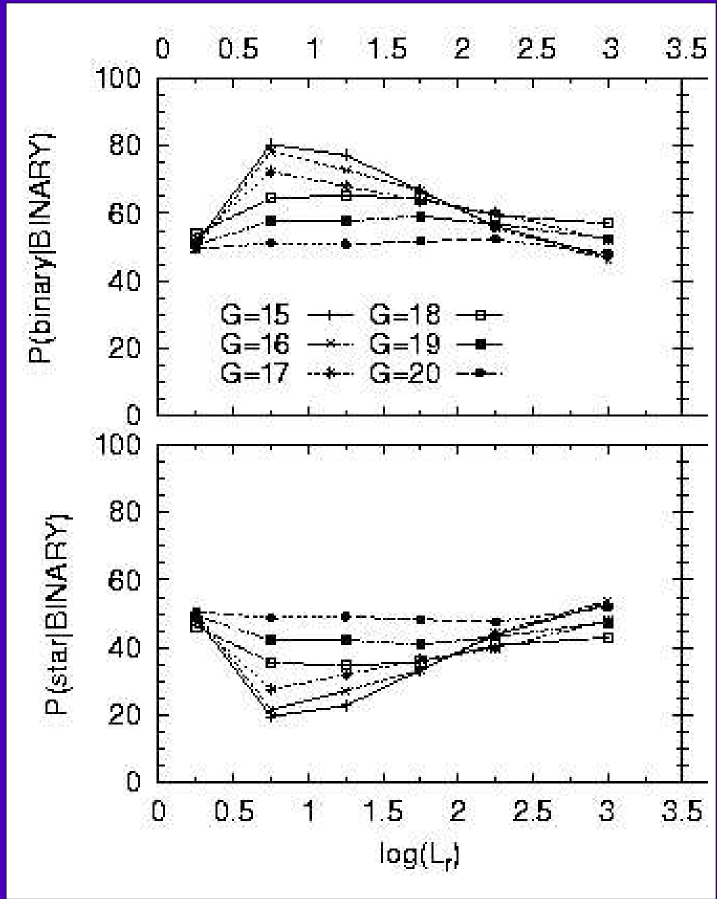
A_V

Bootstrap uncertainty estimates using neural networks (G=15)

Willemsen, Bailer-Jones & Kaempf (2004)

Detection of unresolved binaries

- single star/binary classification
- IMF: Kroupa (2001) + more RGs
- composite MBP photometry
- P(BINARY)= 50% *(prior)*
- Support Vector Machine Classifier



performance is a strong function of component luminosity ratio (L_r) and magnitude (G)

Willemsen, Kaempf & Bailer-Jones (2004)
see poster 4.17

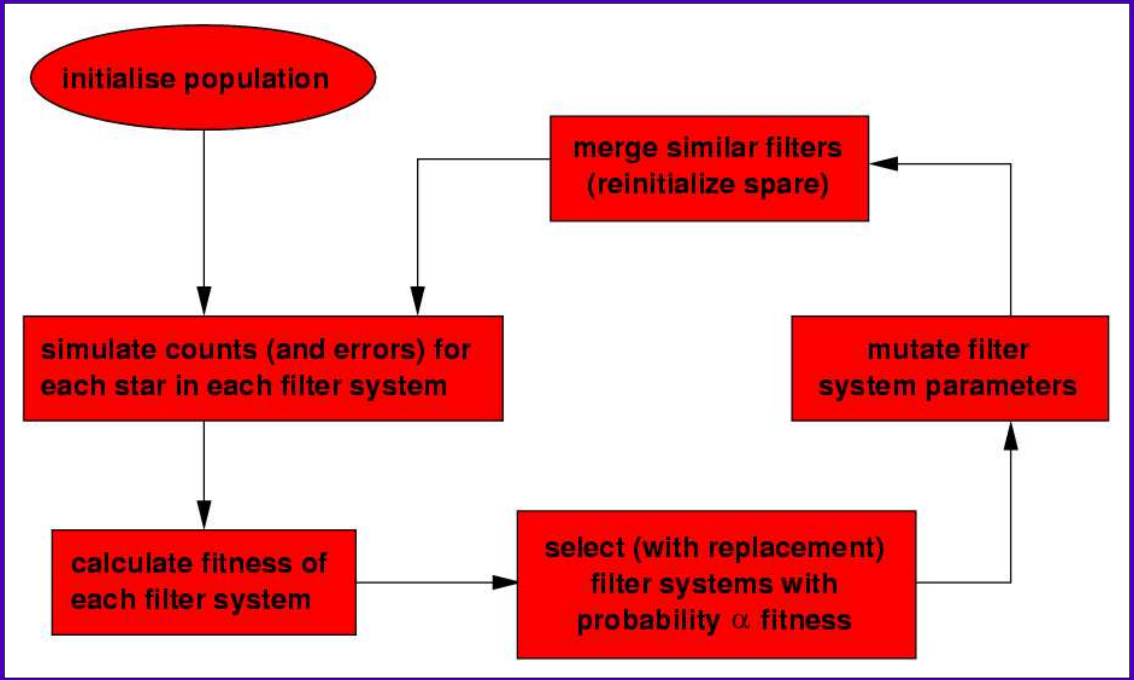
Issues

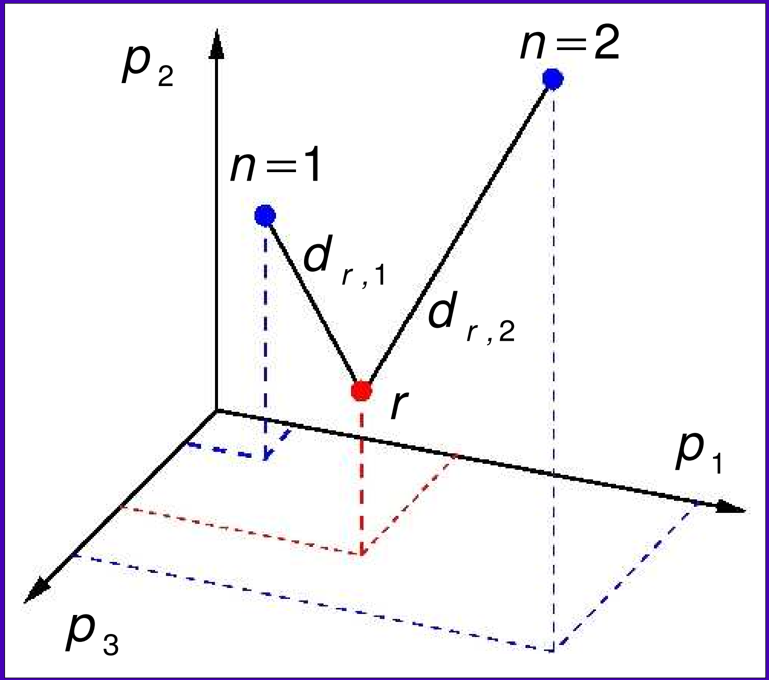
- model dependent physical (supervised) scheme
  - additional empirical (unsupervised) scheme?
- calibration of models
  - e.g. synthetic spectra + Gaia observations of known stars
- model training
  - regularization
- problem of “strong” vs. “weak” APs
  - systematic/correlated errors
- dealing with AP degeneracy
  - data space partioning?
- heterogeneous data
  - optimal combination of BBP, MBP, RVS etc.

Optimization of Gaia filter systems

Coryn Bailer-Jones
Classification with Gaia
Paris, 4-7 October 2004

- optimize the input data $\Rightarrow$ optimize photometric system (PS) for doing APE
- parametrize PS
- define figure-of-merit (FoM) of PS performance
- optimize FoM with respect to PS parameters (evolutionary algorithm)
- see Bailer-Jones (2004, A&A) and poster 4.3





Summary

- Objective: to determine the intrinsic properties of individual objects observed by Gaia
- wide range of objects/APs ⌡ challenging problem
- mutidimensional data analysis methods required
- several algorithmic issues to be addressed
  - degeneracies, training data (calibration)
- numerous open issues
  - discrete source classification, galaxy classification, treatment of peculiar/rare stars, empirical (unsupervised) classification, optimal combination of data (MBP, BBP, RVS, parallaxes etc.)
- several relevant posters (2.6, 2.10, 4.8, 4.9, 4.13, 4.17, 5.16, 5.19)