Classification and Astrophysical Parameter Estimation from Gaia: Design and First Results

C. Tiede, J. Holmberg, C.A.L. Bailer-Jones

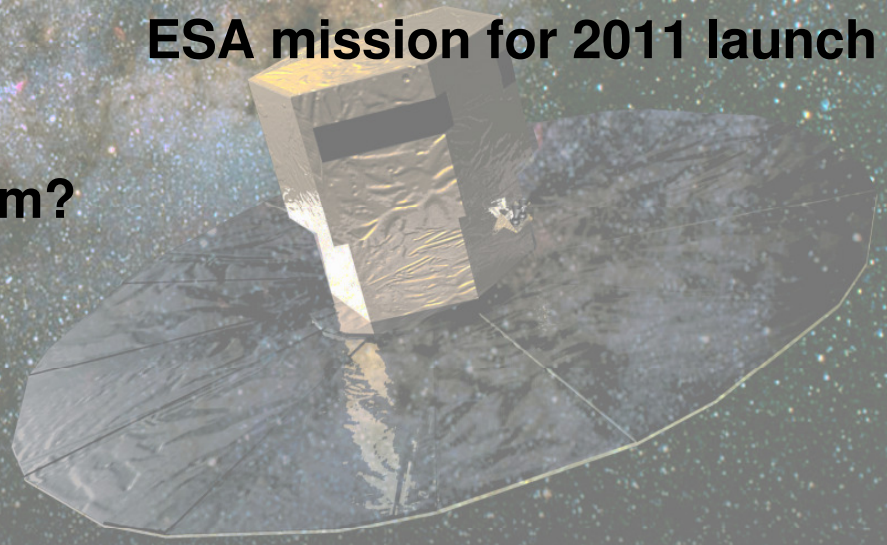
ADASS 2006
Tucson, AZ
October 2006

Gaia in a nutshell

Astrometry
Spectroscopy
Photometry

Entire sky to G=20 (V=20-22),
Up to 100 times over 5 years,
About 1 billion stars, 1 million quasars, few million galaxies

ESA mission for 2011 launch

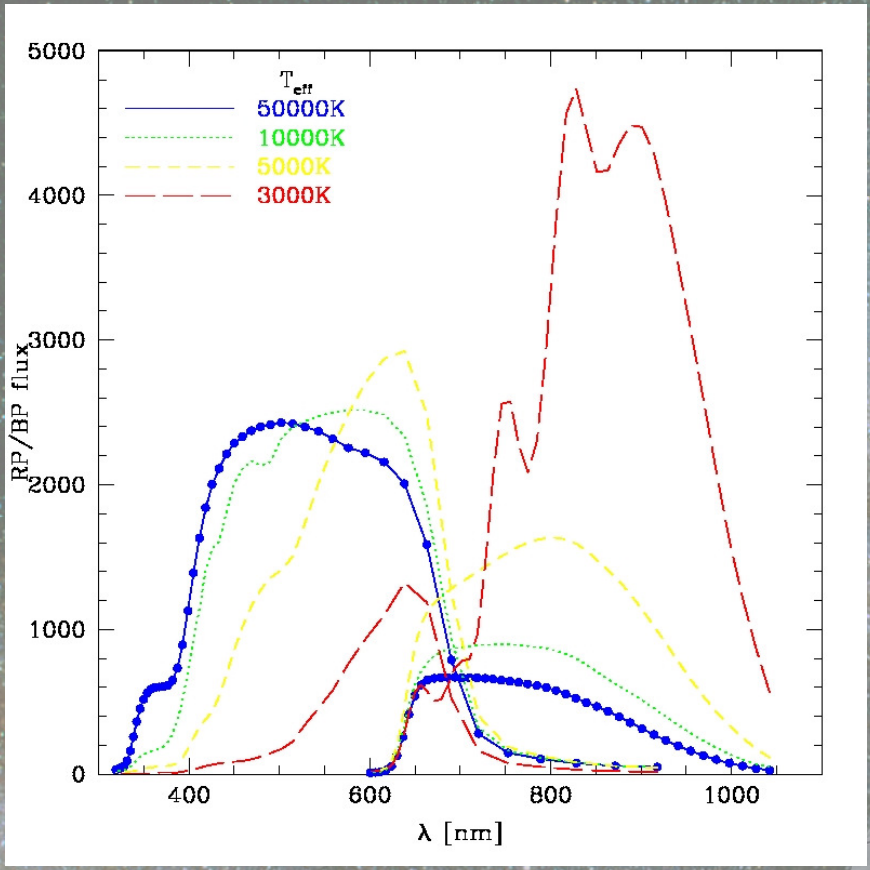


How and when did the Galaxy form?
What is the Galaxy made of?
Substructure in disk and halo
Star formation history
Distribution of dark matter

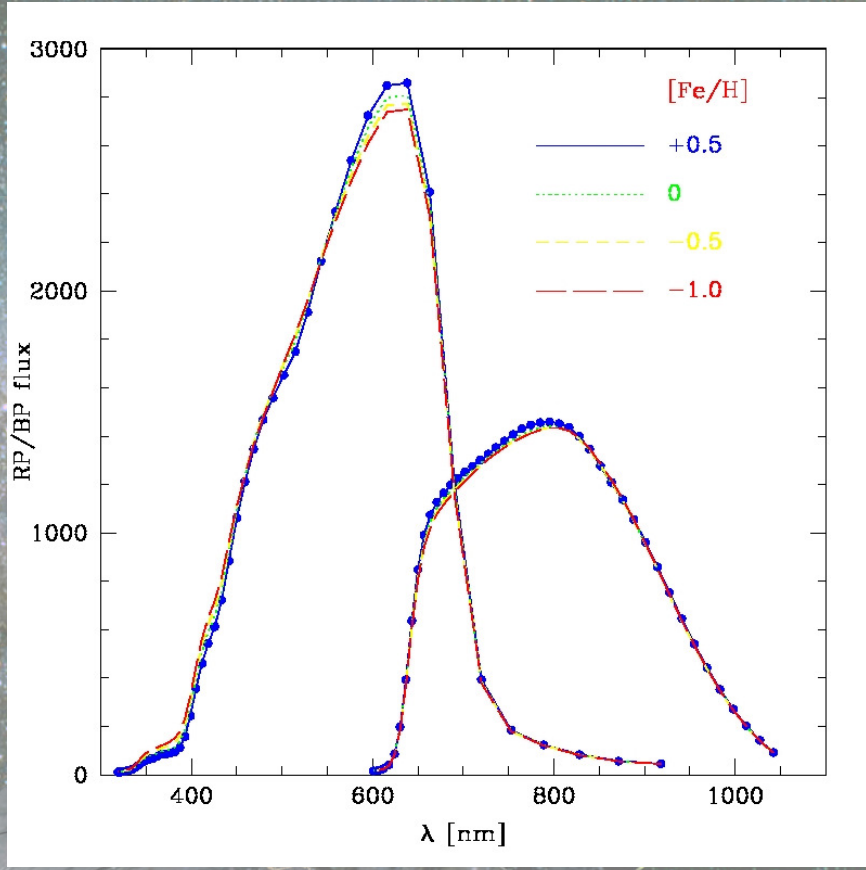
Limitations & Requirements

- **Red and blue photometric data (RP/BP) due to**
  - **Low spectral resolution**
    - **due to low dispersion of prisms**
    - **from broad PSF of optics**
  - **Low signal-to-noise ratio (for faint stars)**
- **Discrete classification of observed objects**
  - **including identification of new types of objects**
- **Determination of intrinsic astrophysical parameters**
  - **Parameter degeneracy**
  - **Problem of “weak” vs. “strong” parameters**

Limitations & Requirements



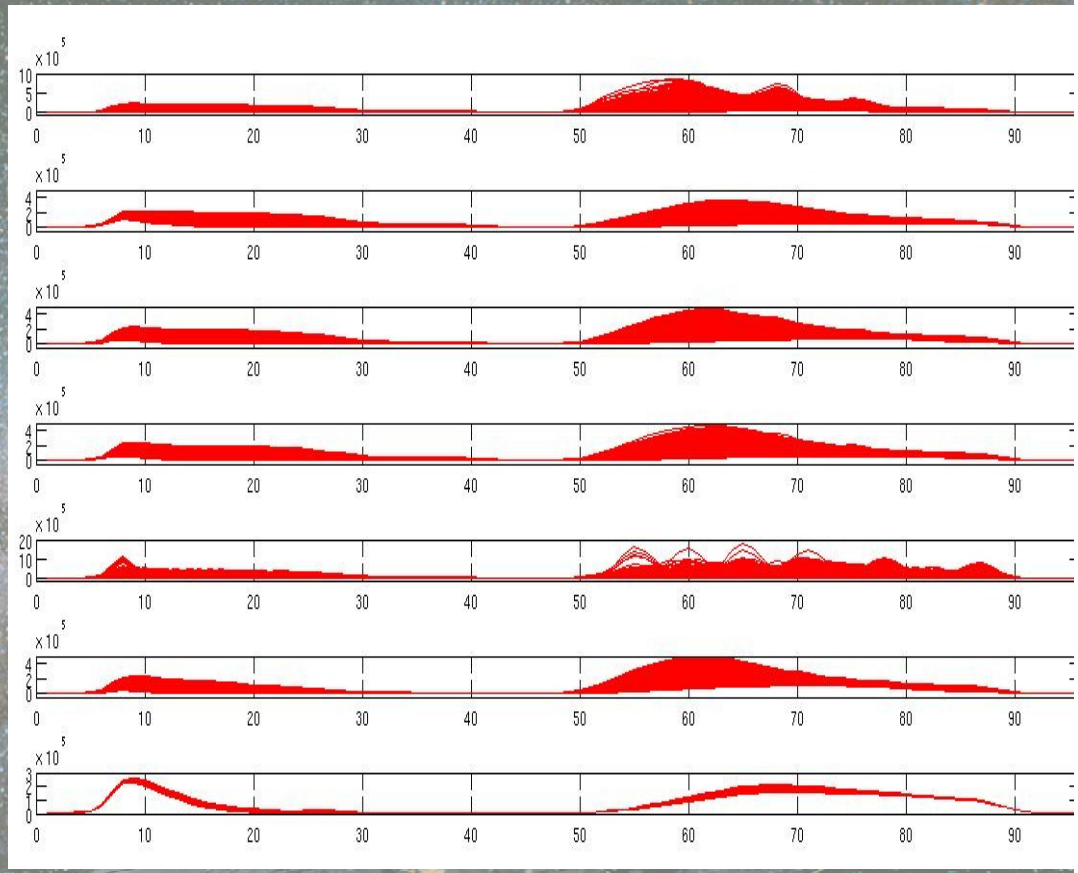
Many variations in RP/BP
→ Strong parameter



Few variations in RP/BP
→ Weak parameter

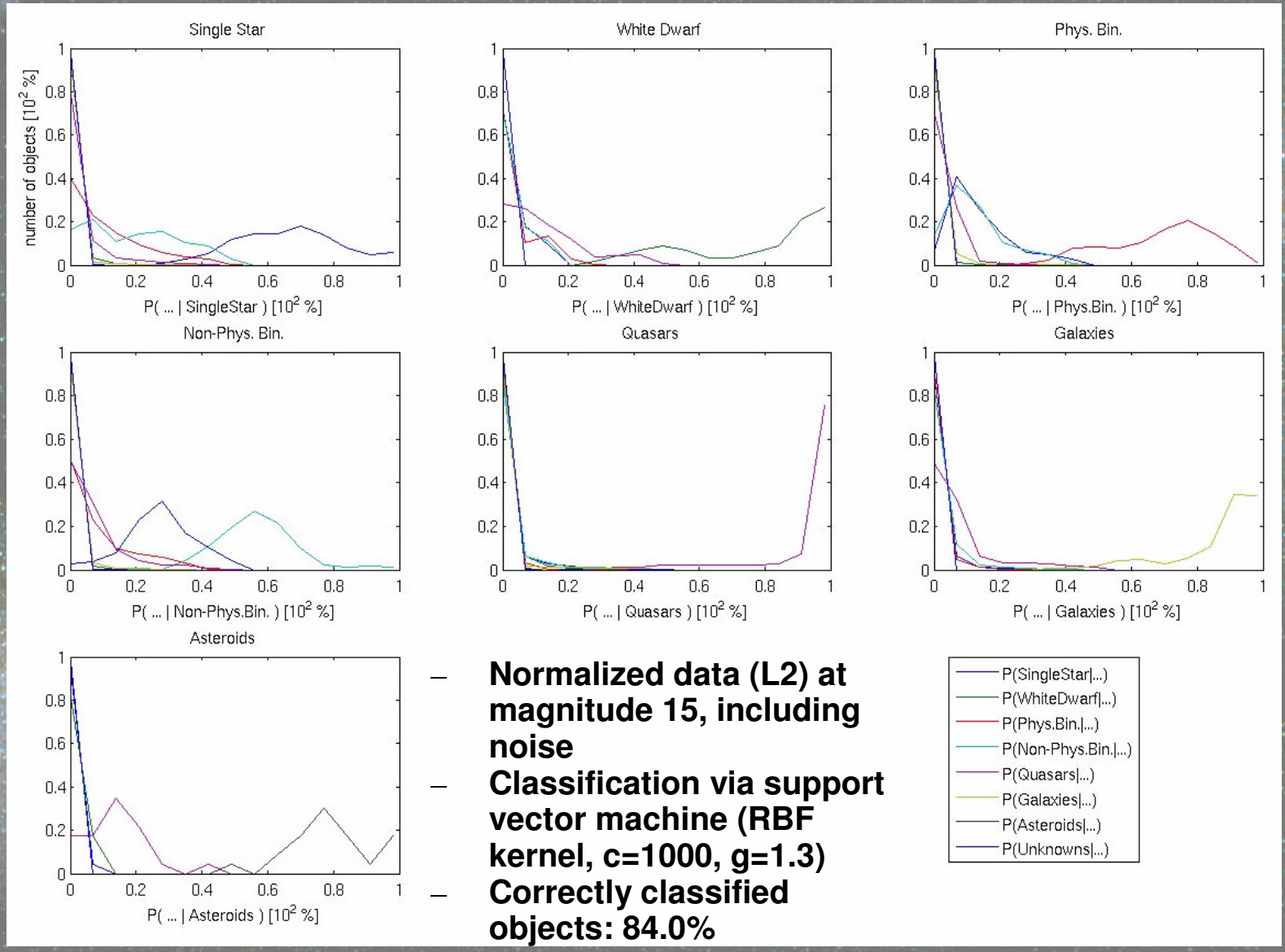
Discrete Source Classification (DSC)

- **Classification into**
  - **single stars**
  - **White Dwarfs**
  - **physical binaries**
  - **non-physical binaries**
  - **quasars**
  - **galaxies**
  - **asteroids and**
  - **+ unknown objects**



Estimation of label + corresponding class probabilities

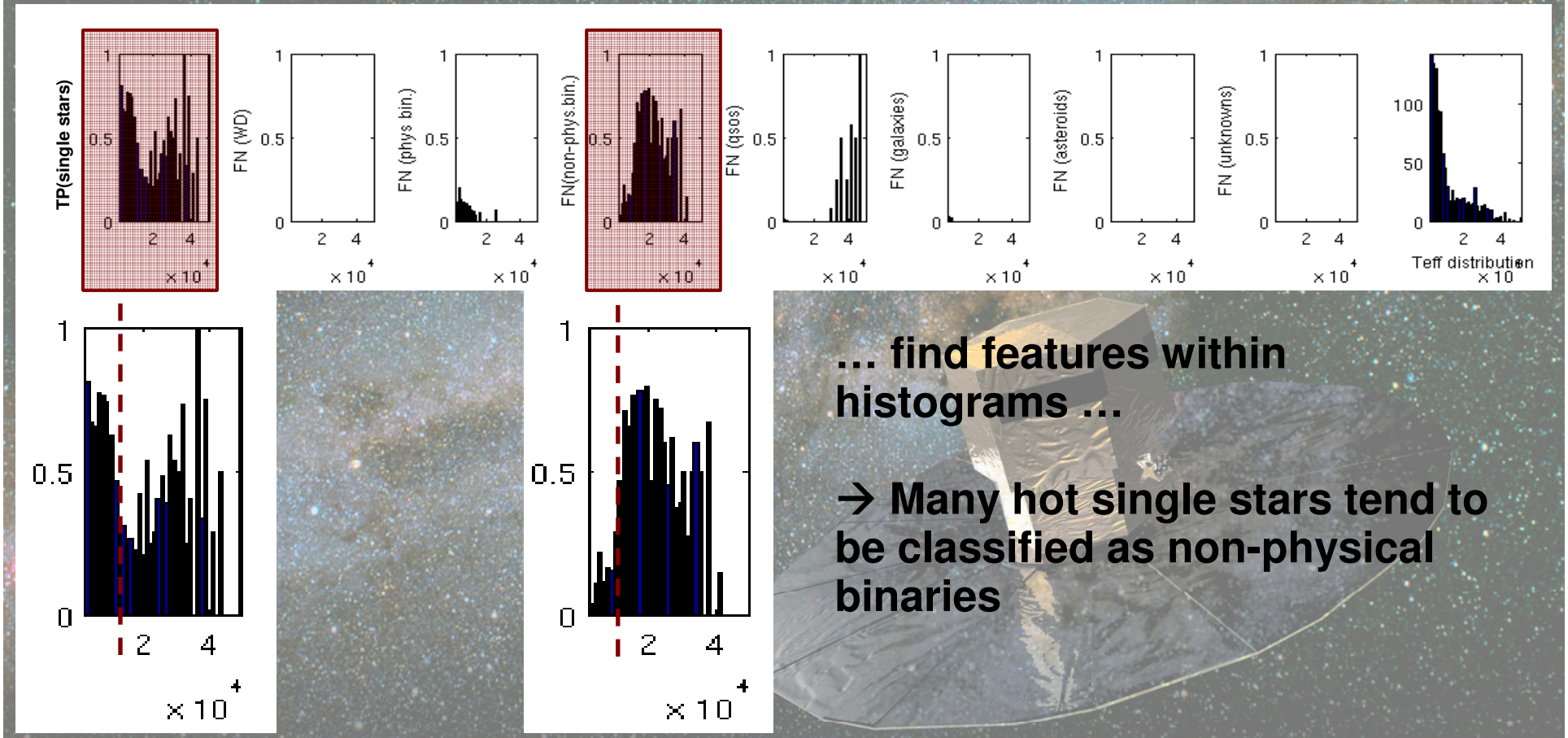
Class Probability Estimates (DSC)



- **Normalized data (L2) at magnitude 15, including noise**
- **Classification via support vector machine (RBF kernel, c=1000, g=1.3)**
- **Correctly classified objects: 84.0%**

True Positives and False Negatives in relation to Astrophysical Parameters

→ **For single stars: Plots in relation to T_{eff}, log g, [Fe/H], Av**

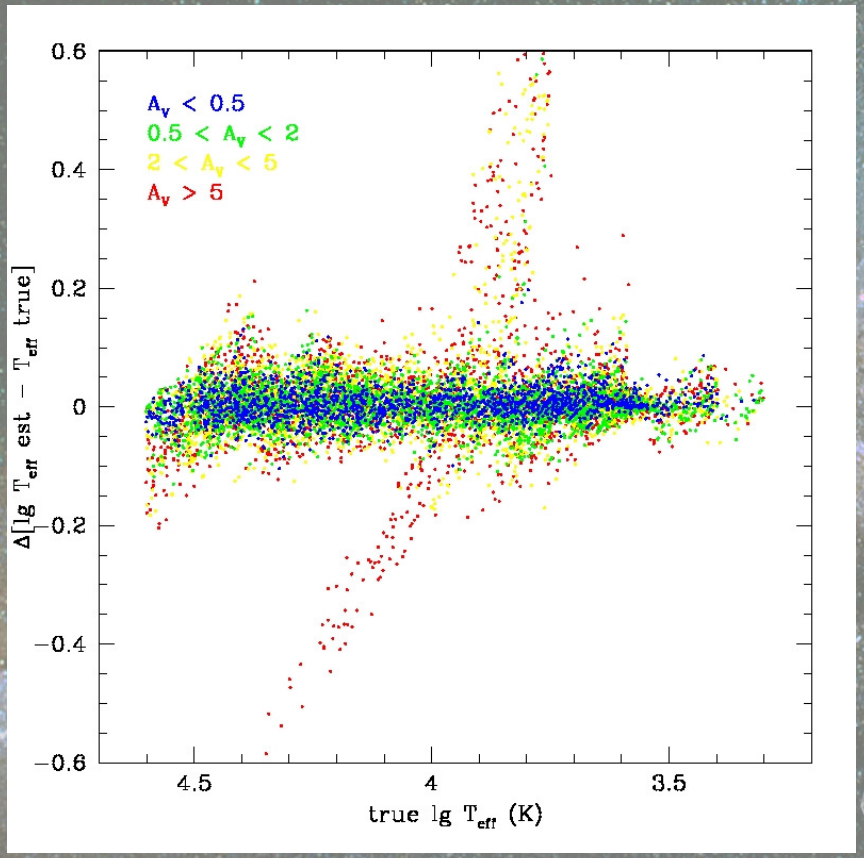


… find features within histograms …

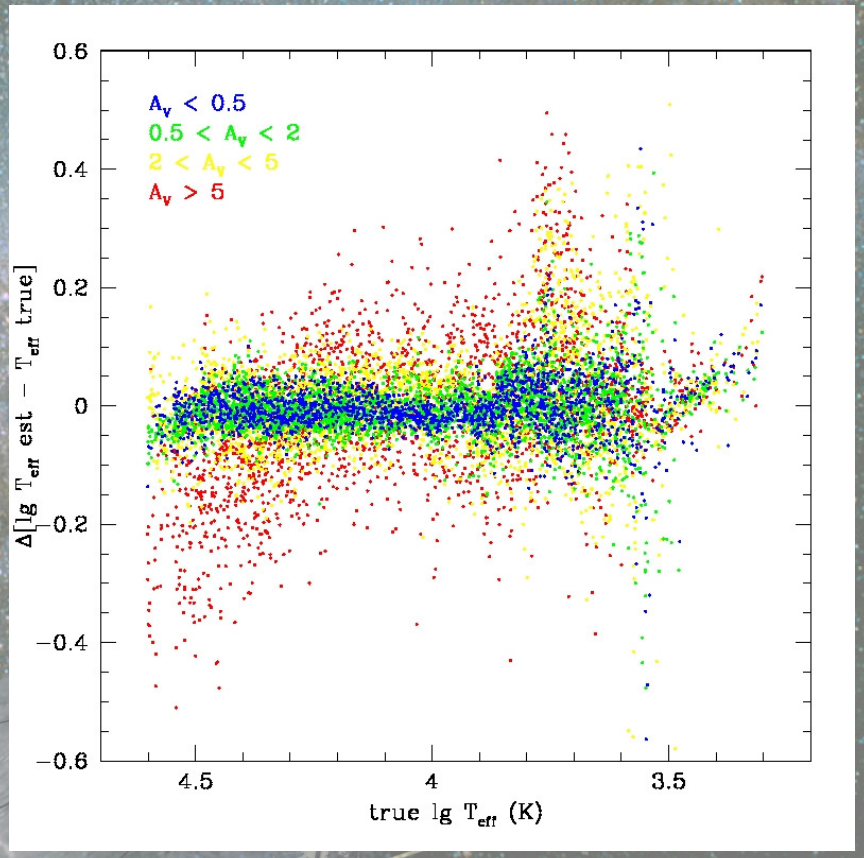
→ Many hot single stars tend to be classified as non-physical binaries

Continuous Parameter Estimation (GSP-phot)

Estimation of T_{eff} (strong parameter)



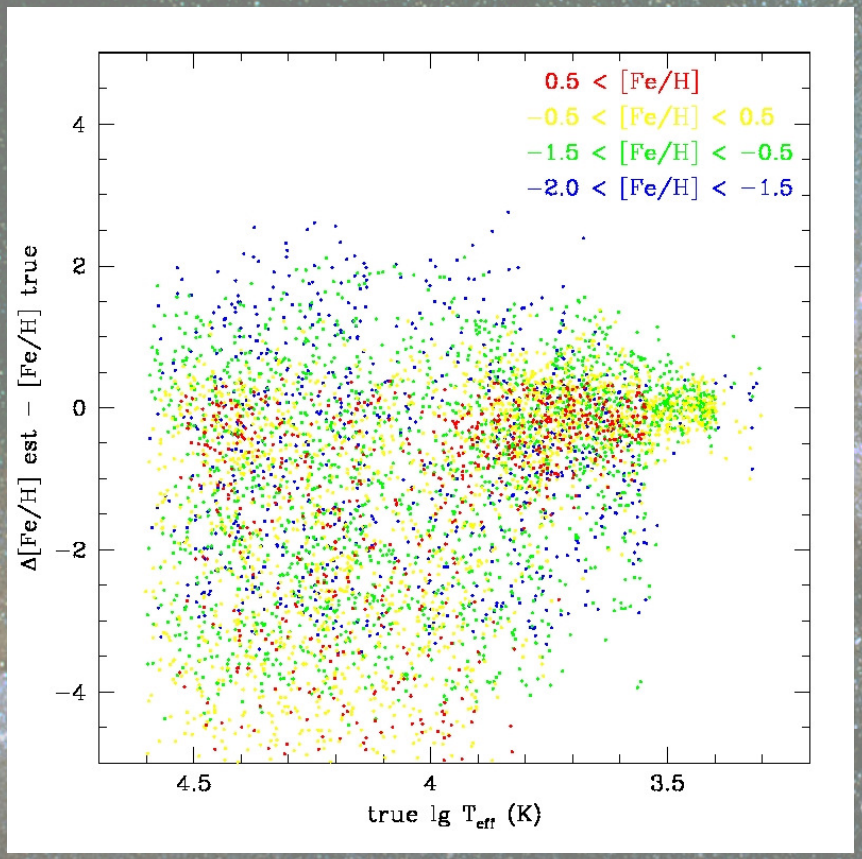
Nearest neighbor approach



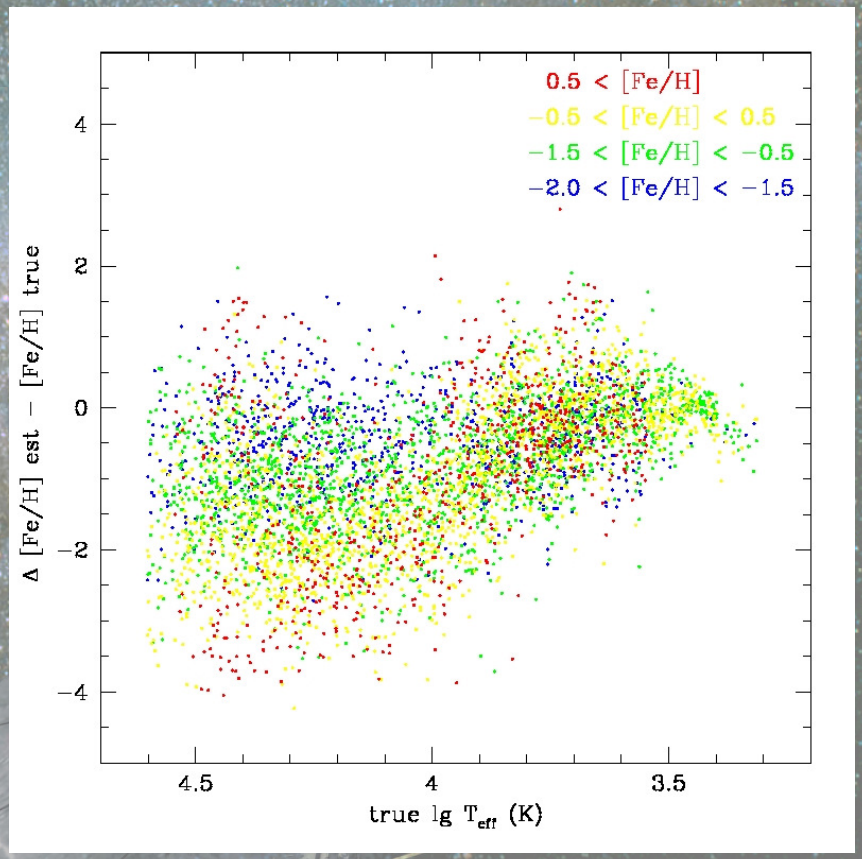
Support vector machine approach

Continuous Parameter Estimation (GSP-phot)

Estimation of [Fe/H] (weak parameter)



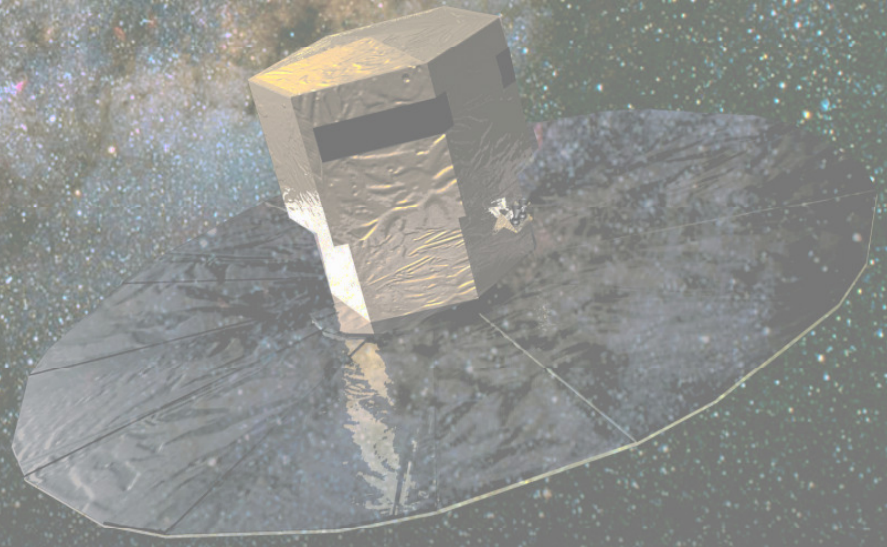
Nearest neighbor approach



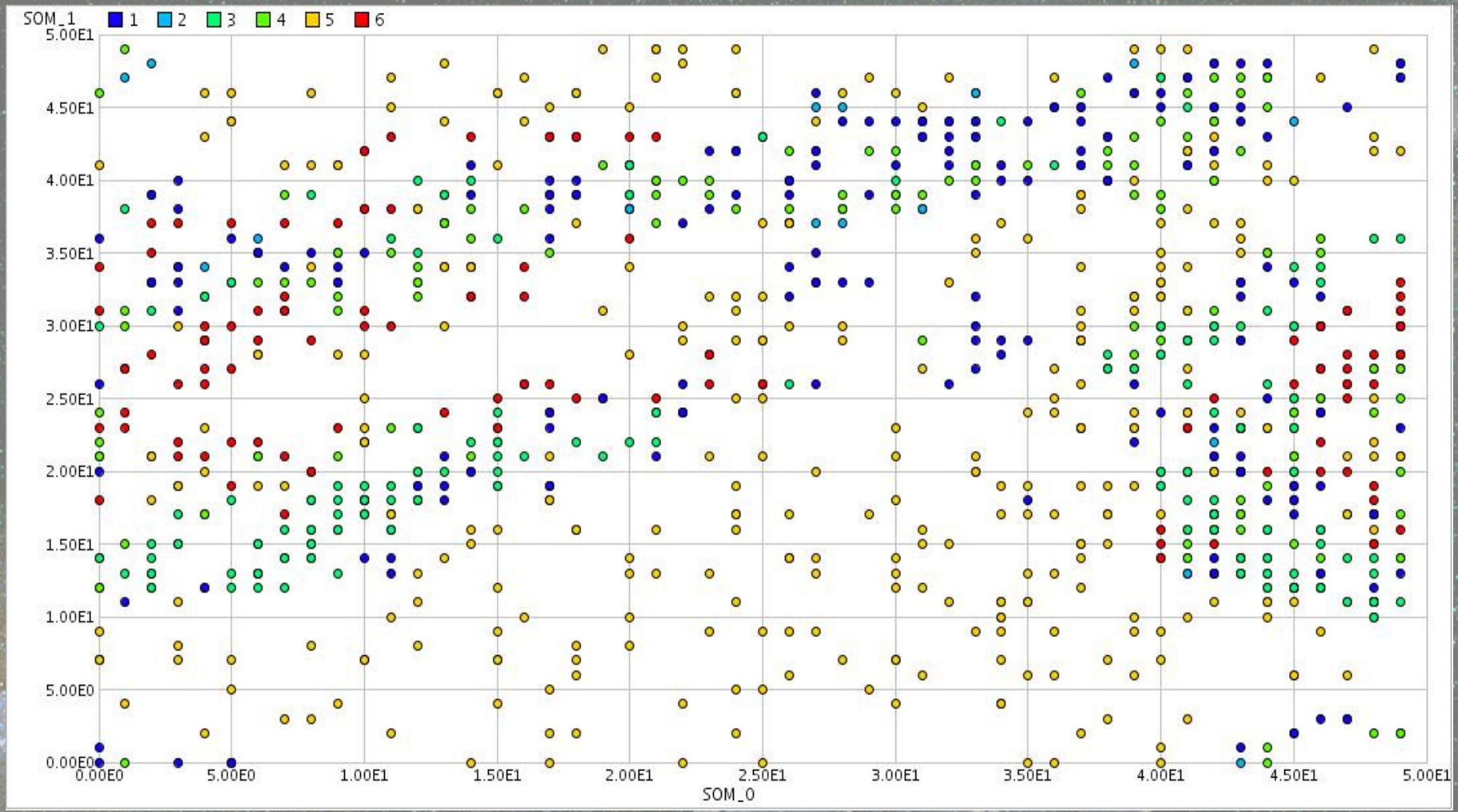
Support vector machine approach

Summary and Future

- **RP/BP data quality**
- **DSC**
  - **Good accuracy for quasars, galaxies, White dwarfs**
  - **Difficulties to distinguish between single stars, physical and non-physical binaries (FP rate ~ 20%)**
  - **→Implementation of unsupervised methods**



Summary and Future



Single stars | WD | Phys. bin | Non-Phys. bin

quasars | galaxies

Summary and Future

- **RP/BP data quality**
- **DSC**
  - **Good accuracy for quasars, galaxies, White dwarfs**
  - **Difficulties to distinguish between single stars, physical and non-physical binaries (FP rate ~ 20%)**

  →Implementation of unsupervised methods
- **GSP-phot**
  - **Good accuracy for strong parameters (e.g. T_{eff})**
  - **Difficulties to estimate weak parameters (e.g. [Fe/H])**