Classification of light curves derived from COROT exoplanet data

Jonas Debosscher

8th Corot week

Toulouse, 23-27 May 2005

Outline

- HIPPARCOS-selection used as test-database
- Application of 2 different classification methods to HIPPARCOS-selection
- Results
- Conclusions
- Future

HIPPARCOS dataset

- Selection made by independent party (Rafa Garrido)
- 1076 HIPPARCOS lightcurves, all kinds of stars
- Only 440 of these were classified earlier by the HIPPARCOS team (visual inspection)
- → Purpose: classify this set using only the lightcurves (+ colours in future)

Classification method:

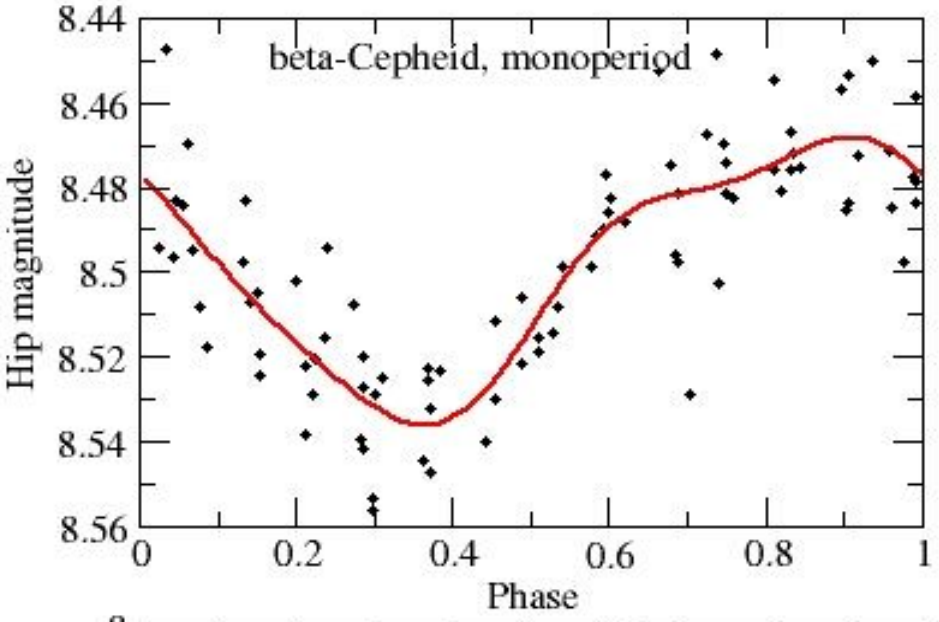
- Pre-processing of lightcurves: remove bad measurements
- Reject ‘non-variable’ stars (given the scatter + measurement errors)
- Determination of relevant parameters describing the lightcurves:
  - Period analysis with PDM (Phase-Dispersion-Minimisation)
  - Make Phase-diagram with period
  - Fit of the phase-diagram using 3 harmonics
  - Determine amplitude of dominant mode and use this to normalise the phase-diagram

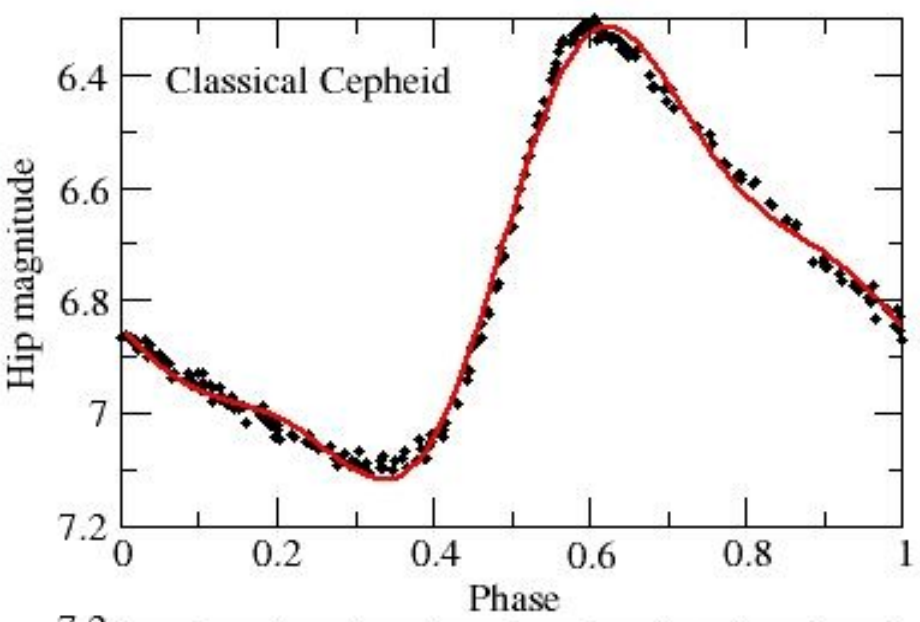
    → Parameters solely describing shape of lightcurve

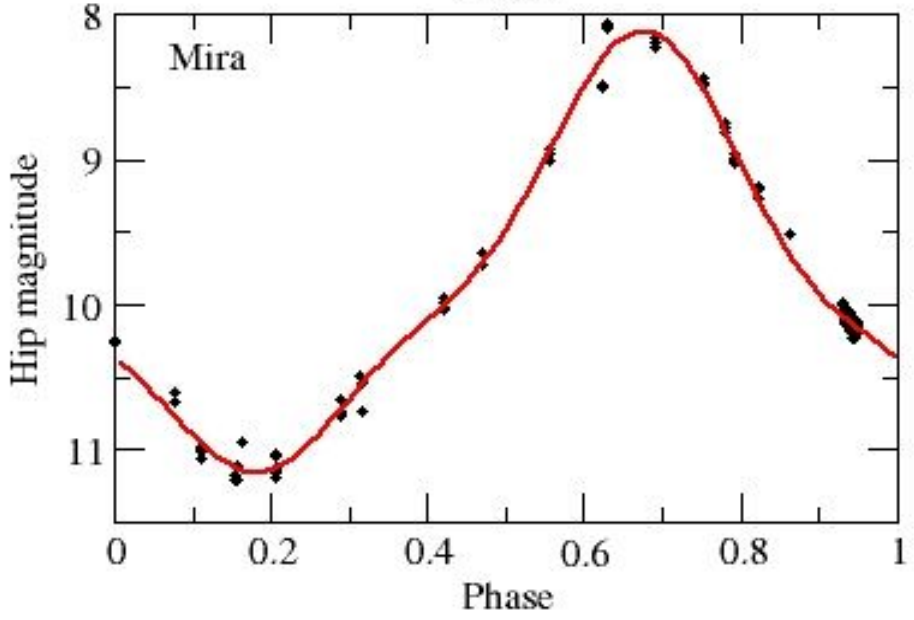
‘Supervised’ classification:

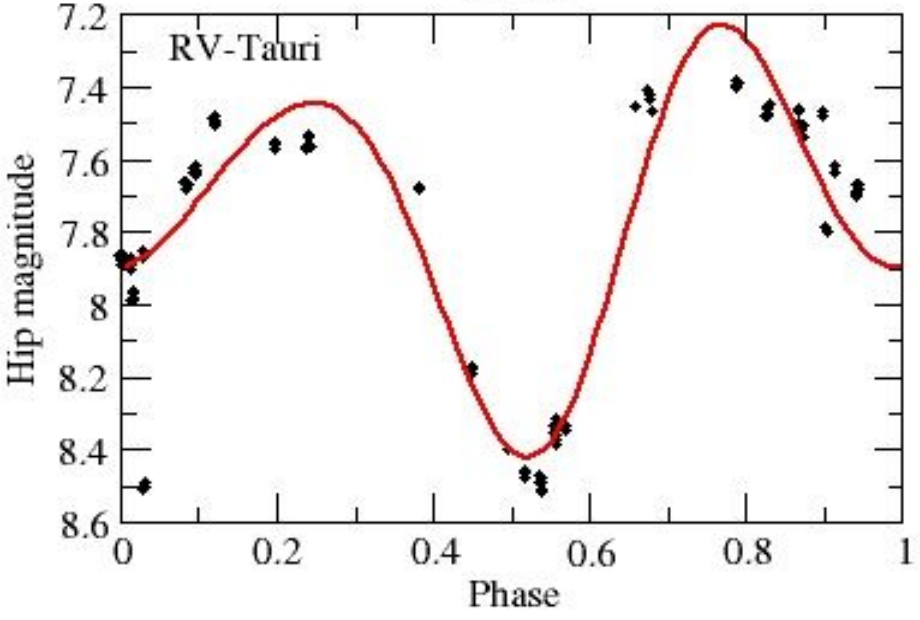
- Classify using lightcurve shape-parameters + amplitude and main period (good description of the lightcurves, except for some types of eclipsing binaries)
- Determine definition classes: use one or more well known HIPPARCOS member stars per class (stars not present in our dataset)
- Calculate the abovementioned parameters for both the definition stars and the objects in the dataset
- Compare every object with the definition stars by calculating distances in multi-dim. parameterspace (for the moment: Euclidian distance)
- Class assignment of star determined by nearest definition star

Some definition stars: phaseplot + harmonic fit

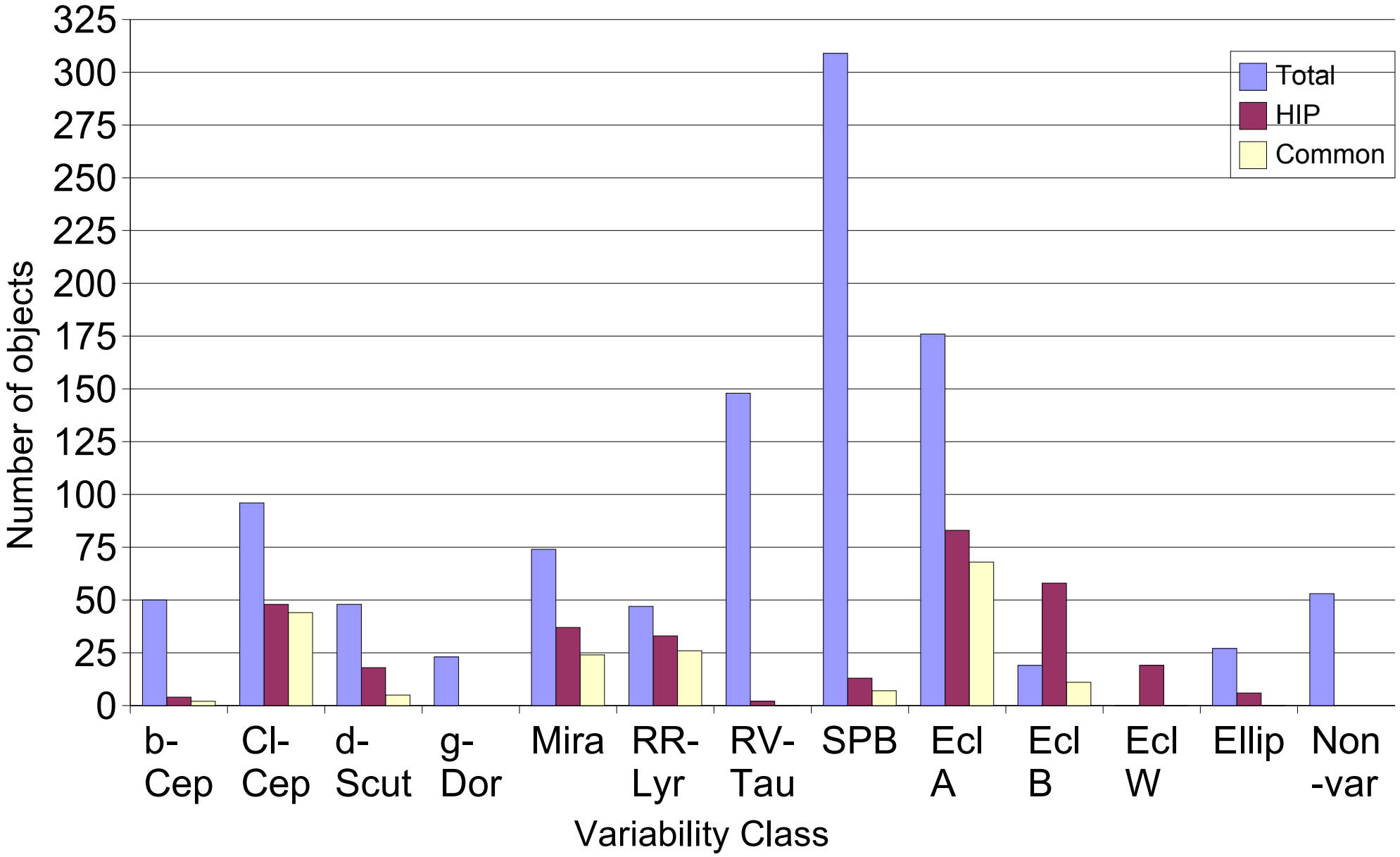




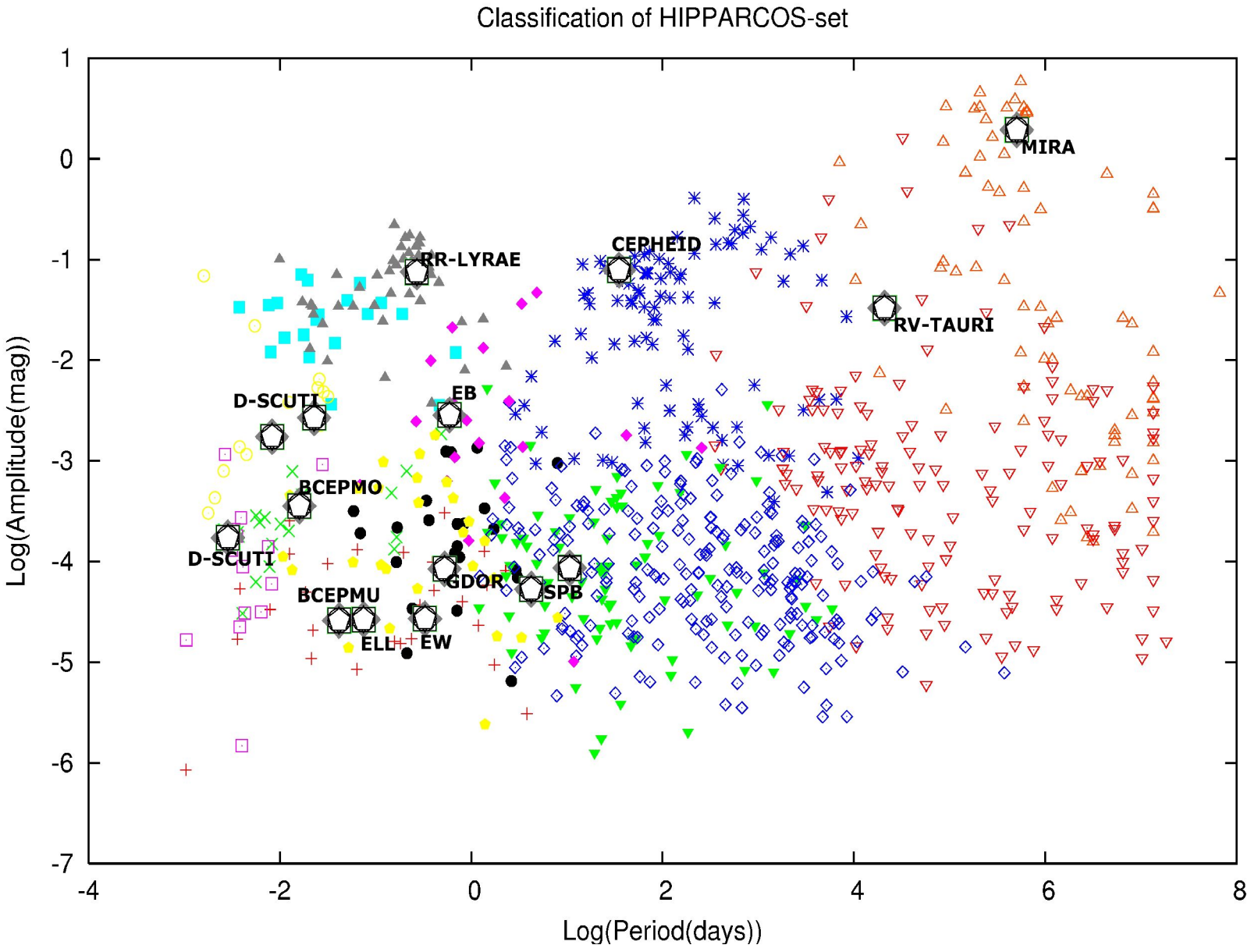




Supervised classification results (using 12 classes) compared with the earlier and incomplete HIPPARCOS classification



Representation in Period-Amplitude plane (2D-projection of multi-dim. parameter space):



'Unsupervised' classification

- Autoclass program (*Cheeseman,1996*):
  - Uses Bayesian statistics to find patterns (classes) in large multi-dimensional parameter spaces
  - Makes no use of definition classes, only the raw input data (fixed number of descriptive parameters for each object)
  - New classes can be found
- Autoclass applied on HIPPARCOS-dataset:
  - → 15 classes found:
    - Good correspondence with supervised results for a few classes (Cepheids, Mira's)
    - Best (smallest) classes contain stars with very similar lightcurves
    - Other (large) classes appear more fuzzy (mixed character)

Conclusions:

- Supervised method shows the clearest results
- Promising results with unsupervised method (Autoclass)
- Use of 3 harmonics OK for fitting pulsating stars, not for some eclipsing binaries with steep lightcurves
  - → For the moment: pre-selection with different method
- Improvements to be made in 2 domains:
  - More accurate and general description of lightcurves
  - Refinement of supervised classification method

Future

- use more/better parameters for describing the lightcurves
- Supervised method: more general distance measure + more class-definition stars
- Apply both methods to OGLE-database
- Fine-tuning of the method is database-dependent:
  - → Concentrate on COROT exoplanet data:
    - Use simulated COROT lightcurves for stars in the two fixed fields
    - Simulated lightcurves of all known class members as class-definitions for supervised method