Analysis of the GAIA Light Curves

Laurent Eyer

Observatoire de Genève

(laurent.eyer@obs.unige.ch)

The problem

GAIA: one billion photometric time series spanning 5 years

$$\text{Time}_i^{(j)}, \text{Magnitude}_i^{(j)}, \epsilon_i^{(j)}$$

$i = 1, \ldots, n$, $j = G, BBP, MBP$, $n \simeq 80$ (40-220) for G and BBP and $n \simeq 200$ (120-380) for MBP (for transit!)

- **"Constant" objects**: Two quantities: mean magnitude and precision (for each filter, radial velocities)
- **"Variable" objects**: variety of behaviours (Parameters of a Fourier decomposition, etc. . .)
  Conservative Forecast (Eyer & Cuypers 2000):
  Number of variable stars: $\simeq$ **18 million**
  Periodic variable stars: $\simeq$ **5 million**

Known classes: a zoo (see 3^d oral presentation)

- Stars:

  Pulsating stars: α Cyg, β Cep, Cep, W Vir, δ Sct, γ Dor, L, M, PV Tel, RR Lyrae, SARV, SPB, SR, SX Phe

  Variability induced by rotation: α CVn, BY Dra, ELL, FK Com, SX Ari

  Eclipsing: EA, EB, EW

  Eruptive stars: γ Cas, RC Bor, RS CVn, UV Ceti, S Dor, WR

  Cataclysmic stars: Supernovae, Novae, Z And

  Note: a star can be a member of several classes!

- Microlensing events
- QSOs
- γ ray bursts (Totani & Panaitescu 2002)
- Asteroids (δmag = 0.1 - 0.7!)

Problems to be resolved

How to:

- Detect variability
- Describe variability
- Classify (Globally) or extract

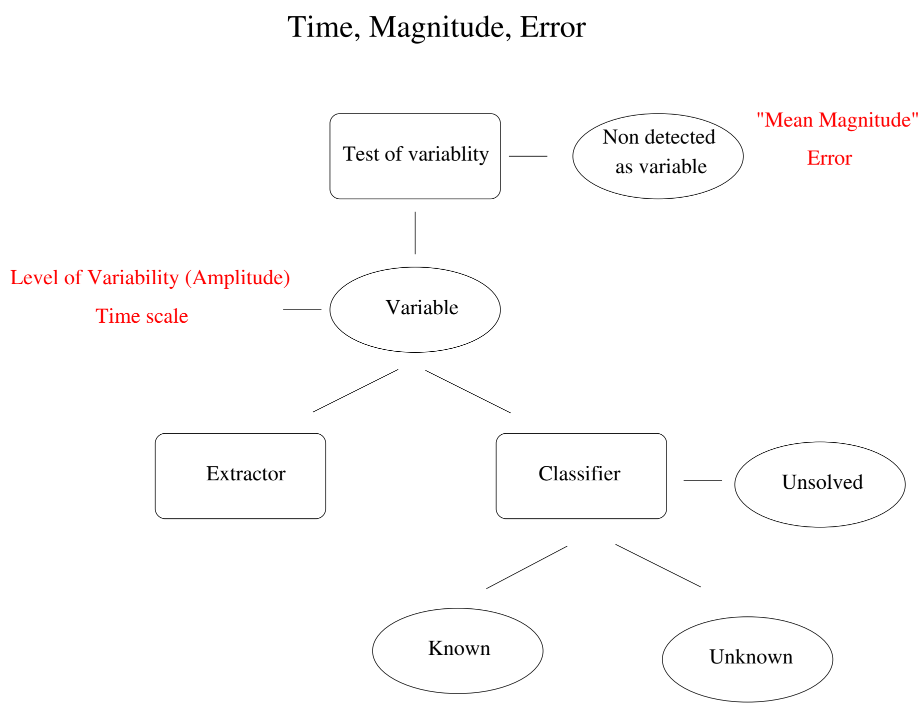
Weakeness/ Strength

- Sparse sampling (variable number of measurements and sampling)
- Precision function of magnitude (and also aging)
- Number of objects (false alarms)

On the other hand:

- High data quality (G band)
- Filter system (”simultaneous”)
- Two ”independent” telescopes (three sets of magnitudes)

Organigram of tasks



Output

- variability class
- variability parameters
- light curves

Non variable $\rightarrow$ photometric calibration

Progressive delivery (not to wait 2018)

Coordination with Earth observations

Other surveys (on which I worked)

- **Hipparcos:** All sky, 3.3 years, 118'000 stars, ~11'500 variables, similarities with GAIA
- **OGLE:**
  - OGLE-II: 4 years, $\sim 40 \times 10^6$ objects, $\sim$ 250'000 variable objects
  - OGLE-III: soon real time photometry (Alarm system)
- **ASAS:**
  - ASAS 1-2, 3 years, 150'000 stars observed, 3900 variables
  - ASAS-3, 1.3 million stars, 1 year, 3126 variables (1046 eclipsing, 778 regular pulsating, 132 Mira, 1170 others)
  - Alarm system in operation (cataclysmic variables)

Other surveys (exploring low level variability)

- **MOST:** Launched 2003, few targets $\simeq 30$
- **COROT:** Launch 2006
  Astroseismology program
  Extrasolar planet program (variability analysis)
- **Kepler:** Launch 2007
  Detection of extrasolar planets

Input parameters

On what to do the analysis? How to combine filters?

- Magnitude, flux, ($\sqrt{\text{flux}}$)
- Group/average sequences of measurements ?
- G magnitude (most precise, calibration difficult)
- Add filters of the MBP or BBP

Variability tests

Tests: magnitudes, magnitudes and their order, full time series

- Khi square test
- Test of outliers
- Abbe test: $r = \frac{n}{2(n-1)} \frac{\sum_{i=1}^{n-1} (s_{i+1} - s_i)^2}{\sum_{i=1}^{n} (s_i - \overline{s})^2}$
- Welch Stetson variability Index
- Test of trends
- Test on peak height in the spectrum

Test the quality of photometric reduction

Preprocessing of the data

Breaking down the data points into meaningful variables

Definition of parameters (example of constrain: Invariance under translation):

- Moments of the magnitude distribution
- Variability level, amplitudes
- Time scales, Periods, slopes, Fourier decomposition, splines, etc. . .
- "Outliers"

Classification Methods

The human brain is bad at seeing groups in multidimensional (≥ 4) space

- Classical approach: stars—ASAS Pojmanski 2002, QSO—OGLE Eyer 2002
- Neural Network: stars—MACHO Belokurov, V. et al. 2002, stars—ROTSE Wozniak, et al. 2004
- Principal Component Analysis: Kanbur et al. 2002 (for cepheids)
- Bayesian approach: stars—ASAS Eyer & Blake 2002, stars—ROTSE Wozniak et al. 2004

Some difficulties

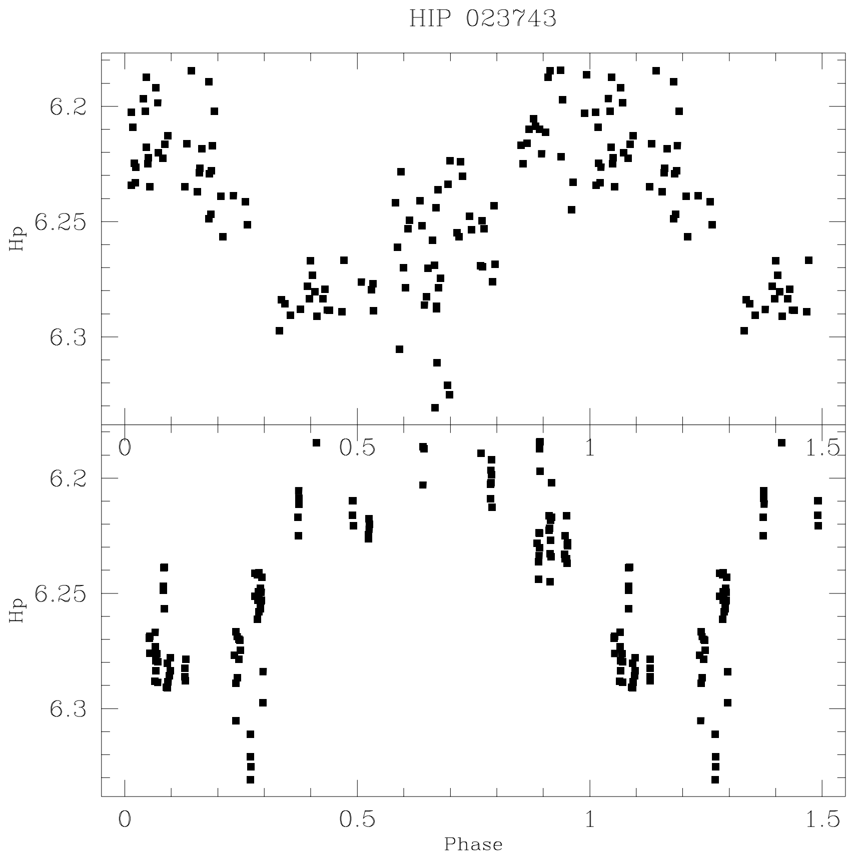
- Epochs of measurement are not identical, number of measurements may be very different (like Hipparcos, but unlike microlensing surveys): difficulty in comparing time series (Koen & Eyer 2002, periodogram analysis, thresholds fixed by permuting data)

- Degeneracy between aliasing problem and non strictly periodic variable phenomena

- GAIA will be rather unique. Microlensing surveys constantly improve (upgrade 4-5 years): no need to spend much time to extract everything from the data

Difficulties (example 1): aliasing

Which period is correct? (top: 3 days, bottom: 81 days)



Difficulties (example 2): Rate of correct detection

To study and understand the consequence of the sampling law

Give a sinusoidal signal with Period P
Sample it with GAIA sampling law
Search period $\longrightarrow$ P'

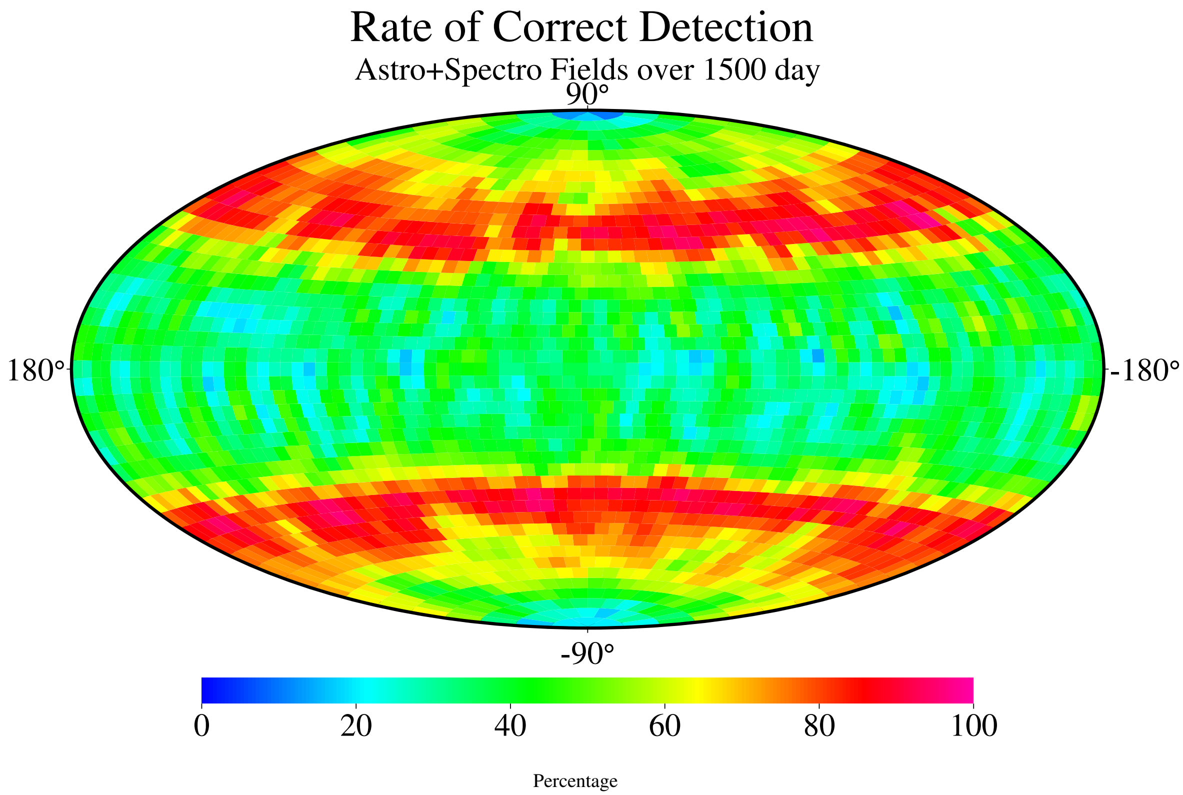
Eyer & Mignard computed over 1'000'000 Fourier transforms

- What are the periods, that are most difficult to recover
- where the aliases are spreading

Difficulties (example 2): Rate of correct detection (see 2^d talk)

For S/N: 0.75 , period: 20 days, the rate varies from $\sim 0\%$ to $\sim 100\%$

(example with old scanning law)



Global approach: ASAS 1-2

- ASAS:
  PI: G.Pojmanski
  Location: Las Campanas
  Telescope: 135 mm telephotolens
  Limiting Mag: I=13
  Surface: 50 field of 2x3 deg^2
  Nb of stars: 150'000 stars, 3900 variables

- Bayesian classifier: Autoclass (Cheeseman & Stutz)
  Previous applications: IRAS spectra, ISO spectra, SDSS asteroid colors, Hipparcos variable star data (M_v, $V - I$, Amp, Per, skewness) (not published)

Application to ASAS 1-2

- Period search (Lomb algorithm)
- Fourier decomposition with n harmonics

Parameters on a well behaved sample of 458 stars (beware of overparametrization):

- Period
- Amplitude
- Skewness
- Ratio of amplitudes (A_2/A_1) (phase difference ($\phi_1 - 2\phi_2$))

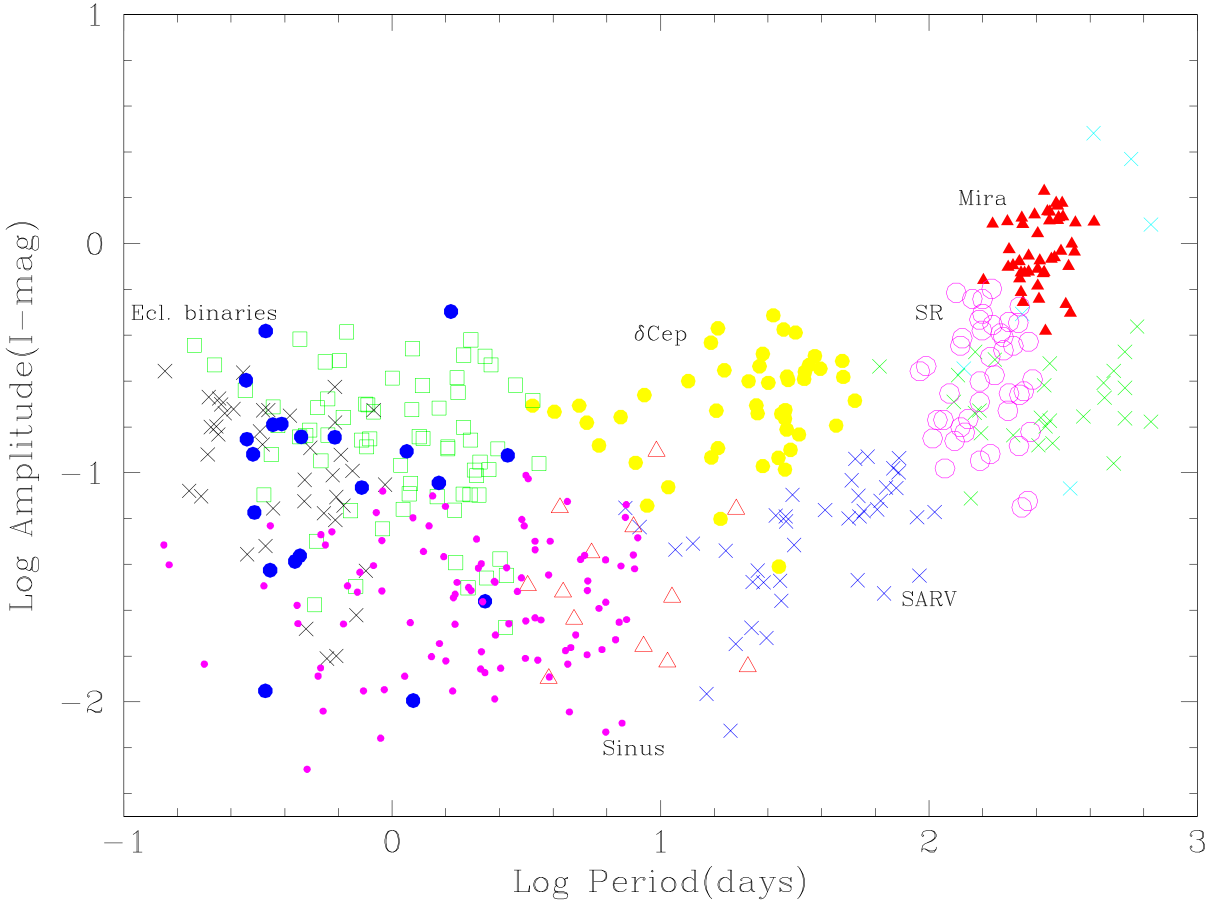
Application to ASAS 1-2

Result of the classification (on 458 stars). Classes defined by Autoclass could be associated to known classes:

- Small amplitude and sinusoidal curves: $\sim$100
- Eclipsing binaries: $\sim$144
- Cepheids: $\sim$48
- SR: $\sim$81
- Mira: $\sim$45
- SARV: $\sim$40

Error level of the classification is 5%

Application to ASAS 1-2

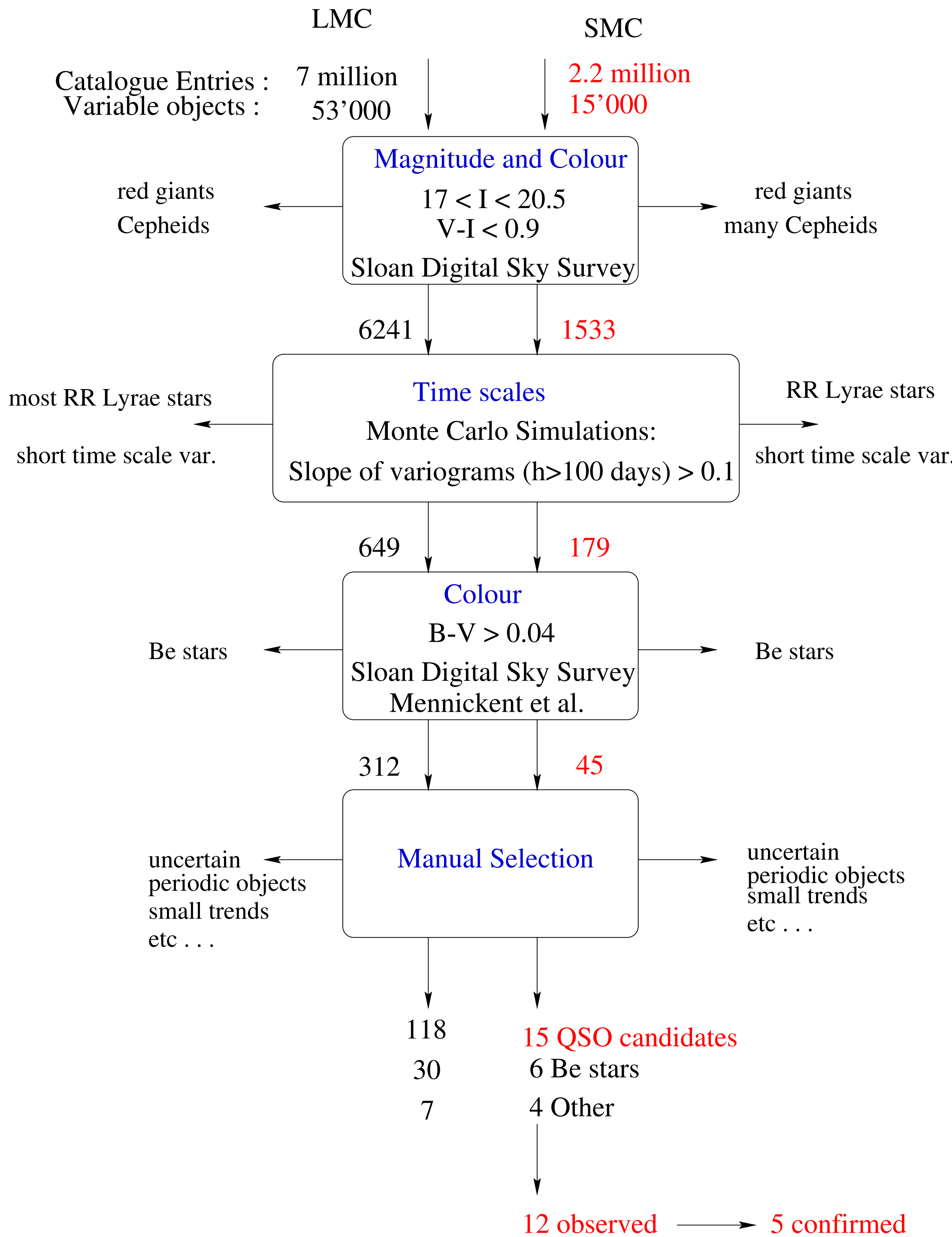


Classical approach (Extractor): Selection of QSOs

Search of QSOs behind the Magellanic clouds in OGLE-II data
Here: Focus on SMC only

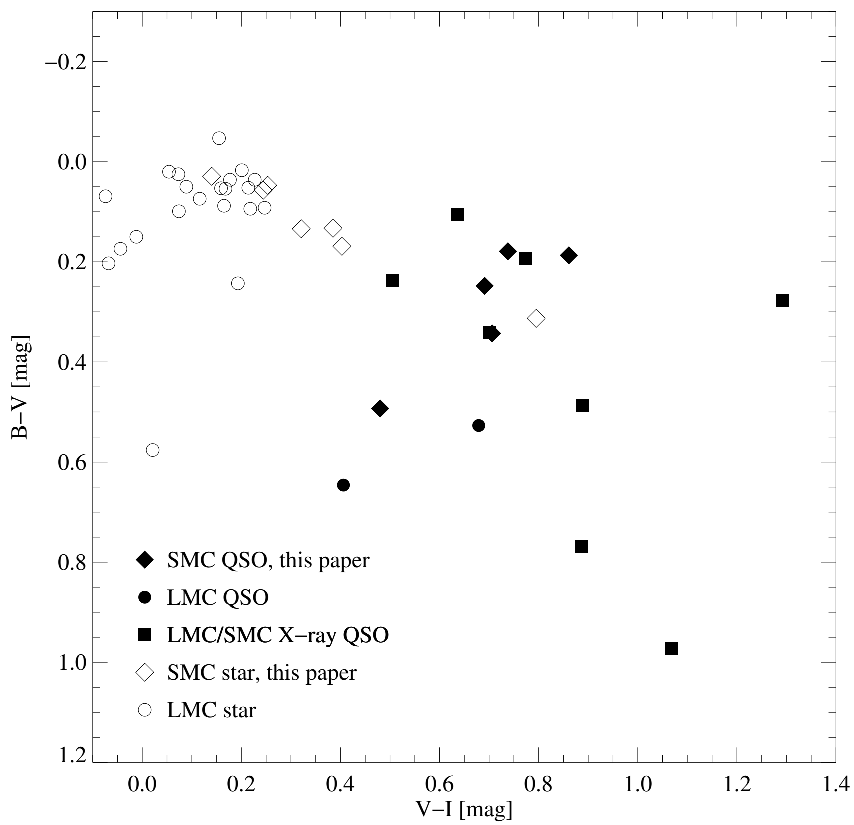
- B, V, I photometry (magnitude and colour cuts)
- Variability

$\rightarrow$ 40% of success


LMC
SMC
Catalogue Entries :
Variable objects :
7 million
53'000
2.2 million
15'000
Magnitude and Colour
17 < I < 20.5
V-I < 0.9
Sloan Digital Sky Survey
red giants
Cepheids
red giants
many Cepheids
6241
1533
Time scales
Monte Carlo Simulations:
Slope of variograms (h>100 days) > 0.1
most RR Lyrae stars
short time scale var.
RR Lyrae stars
short time scale var.
649
179
Colour
B-V > 0.04
Sloan Digital Sky Survey
Mennickent et al.
Be stars
Be stars
312
45
Manual Selection
uncertain
periodic objects
small trends
etc . . .
uncertain
periodic objects
small trends
etc . . .
118
30
7
15 QSO candidates
6 Be stars
4 Other
12 observed
5 confirmed

Selection of QSO. Colour-Colour Diagram

Results from Dobrzycki et al., 2002, will boost up the rate!



What should be done (I)

- What input parameters (formation of groups?)
- Variability and levels of variability
- Time scales and periods:
  - Comparison of period search algorithms
  - Variograms (structure functions), autocorrelation
  - Study of sampling
- Classification method comparison: unsupervised methods (unknown classes), supervised methods, extraction methods

What should be done (II)

Working plan and schedule. First deadline: September 1 2004

What, Who and When ?

- Detection of variability
- Definition of Variable Parameters
- Classification