Building an automated 100 million+ variable star catalogue for Gaia

Berry Holl¹

presented by: Laurent Eyer¹

Nami Mowlavi¹, Dafydd W. Evans², Gisella Clementini³, Jan Cuypers⁴, Alessandro Lanzafame⁵, Joris D. Ridder⁶, Luis Sarro⁷, Diego Ordoñez-Blanco¹, Krzysztof Nienartowicz¹, Jonathan Charnas¹, Leanne Guy¹, Grégory Jévardat de Fombelle¹, Isabelle Lecoeur-Taïbi¹, Lorenzo Rimoldini¹, Maria Süveges¹, François Bouchy⁸¹, Sara Regibo⁶, Mauro López⁹, Jonas Debosscher⁶, Fabio Barblan¹

¹Department of Astronomy, University of Geneva, Versoix, Switzerland
²Institute of Astronomy, University of Cambridge, Cambridge, United Kingdom
³INAF Osservatorio Astronomico di Bologna, Bologna, Italy
⁴Royal Observatory of Belgium, Brussels, Belgium
⁵Dipartimento di Fisica e Astronomia, Universita di Catania, Catania, Italy
⁶Instituut voor Sterrenkunde, KU Leuven, Leuven, Belgium
⁷Departamento de Inteligencia Artificial, UNED, Madrid, Spain
⁸Institut d'Astrophysique de Paris, CNRS, Paris, France

IAU XXIX General Assembly, Hawaii, Division G: Large-Scale Surveys and Variability (DG.6.02)
Gaia challenge

~72 epochs in 5 years

>1,000,000,000 objects, ~10% variable

How to automatically detect, classify and characterise?
Gaia challenge

~72 epochs in 5 years

>1,000,000,000 objects, ~10% variable

Gaia integrated approach:
- Multiple instruments
- Instrument calibration
- Variability detection
- Characterisation
- Classification
- Type-specific modelling
- Iterative training set improvement

How to automatically detect, classify and characterise?
Detect, classify, and characterise variable stars

Gaia FoV: 0.7 deg x 0.7 deg
pixel: 0.059”(AL) x 0.177”(AC)
Detect, classify, and characterise variable stars

Gaia focal plane (106 CCDs)
detection and FOV discrimination

astrometric measurements

photometry (dispersed images)

radial velocity (dispersed images)

BAM = basic angle monitor, WFS = wavefront sensor

WFS1

WFS2

BAM1

BAM2

SM1

SM2

AF1

AF2

AF3

AF4

AF5

AF6

AF7

AF8

AF9

BP

RP

RVS1

RVS2

RVS3

0.42 m

0.93 m

~4.4 sec

~80 sec

Gaia FoV: 0.7 deg x 0.7 deg
pixel: 0.059"(AL) x 0.177"(AC)

Berry Holl (Department of Astronomy, University of Geneva)
Detect, classify, and characterise variable stars

Source signals

Measured signals + noise

Detect, classify, and characterise variable stars

Multiple instruments

Photometry

Spectroscopy

Astrometry

i.e. both photometric and astrometric signal
Multiple instruments

Source signals

Measured signals+noise

signal detection

Detect, classify, and characterise variable stars

Figure adapted from Eyer & Holl, et al. (2013)
Gaia challenge

How to automatically detect, classify and characterise 100 million+ variable stars?

Gaia integrated approach:

- Multiple instruments
- Instrument calibration
- **Variability detection**
- **Characterisation**
- **Classification**
- **Type-specific modelling**
- Iterative training set improvement
Variability detection

- Measured signals + noise
- Variability detection
  - General variability detection
    - Constancy rejection based on p-values of e.g.: $\chi^2$, Abbe, Kurtosis, Skewness, ...
  - Special variability detection
    - Signal matching filters
      - short timescale [10s, 2h]
      - e.g. eclipsing binaries, pulsating stars
      - solar-like activity (spots and flares)
      - small periodic amplitude
      - planet transits
- Select for variable processing (~10%)
- Classify and characterise variable stars

>1,000,000,000 time series
(spectro)photometric bands, radial velocity, ...
Signal characterisation

Measured signals + noise

Variability detection

Characterisation

Classify variable stars

>1,000,000,000 time series
(spectro)photometric bands, radial velocity, ...

General variability detection

Special variability detection

Periodicity / stochastic / ....

Attribute computation

• Periodic model:
  multi-periodic harmonic Fourier
  with polynomial (e.g. trend)

• Many parameters available: i.e. astrometric, astrophysical, radial velocity (for bright stars), etc.

• Various attributes, e.g.:
  absolute magnitude (using parallax),
  BP-RP (colour),
  peak-to-peak amplitude,
  period,
  skewness,
  A2/A1 (ratio of Fourier harm. ampl.), etc.
Classification

Measured signals+noise

Variability detection

Characterisation

Classification

>1,000,000,000 time series
(spectro)photometric bands, radial velocity, ...

General variability detection

Special variability detection

Periodicity / stochastic / ....

Attribute computation

Supervised

Unsupervised

• **Single and multi-stage:**
  Gaussian Mixture, Random Forest, Bayesian network, ...

• **Extractors:**
  transients, microlensing

• **Meta-classifier:** combination of classifiers for best performance.

Identify/confirm new classes
Hierarchical Modal Association Clustering (HMAC)
Type-specific modelling

Measured signals+noise
Variability detection
Characterisation
Classification
Detailed modelling

>1,000,000,000 time series
(spectro)photometric bands, radial velocity, ...

General variability detection
Periodicity / stochastic / ....
Attribute computation

Supervised

Unsupervised

Specific Object Modelling

- RR-Lyrae
- Cepheids
- Eclipsing binaries
- Long period Var.
- Cataclysmic
- BE stars
- Pre main sequence
- Rapid phase
- Microlensing
- AGN
- Planetary transit
- Solar like oscillators
- Flare stars

Berry Holl (Department of Astronomy, University of Geneva)
Gaia challenge

How to automatically detect, classify and characterise 100 million+ variable stars?

**Gaia integrated approach:**
- Multiple instruments
- Instrument calibration
- Variability detection
- Characterisation
- Classification
- Type-specific modelling
- **Iterative training set improvement**
Iterative training set improvement

First iteration:

- Literature
  - start
  - set

- Supervised
  - train

- training set

- Specific Object Modelling

= requires visual inspection/confirmation
Iterative training set improvement

Further iterations:

- Supervised classification
- Semi-supervised training
- Specific Object Modelling
- Literature update
  - New class
  - Add/remove/relabel
- Unsupervised

compare clusters with literature: identify new classes

= requires visual inspection/confirmation

Berry Holl (Department of Astronomy, University of Geneva)
First Gaia data!

28 day Ecliptic Pole Scanning data (summer 2014)

Field-of-view transits during EPSL (equatorial)

Berry Holl (Department of Astronomy, University of Geneva)
First Gaia data!

28 day Ecliptic Pole Scanning data (summer 2014)

Number density of stars up to magnitude 20

Berry Holl (Department of Astronomy, University of Geneva)
First Gaia data!

28 day Ecliptic Pole Scanning data (summer 2014)

Number density of EPSL measurements up to magnitude 20

Berry Holl (Department of Astronomy, University of Geneva)
EPSL data examples

Short period/faint magnitude Cepheids in the Large Magellanic Cloud

Gaia Image of the week:
http://www.cosmos.esa.int/web/gaia/iow_20150528

Credits: ESA/Gaia/DPAC/CU5/DPCI/CU7/INAF-OABo/INAF-OACn Gisella Clementini, Vincenzo Ripepi, Silvio Leccia, Laurent Eyer, Lorenzo Rimoldini, Isabelle Lecoeur-Taibi, Nami Mowlavi, Dafydd Evans, Geneva CU7/DPCG and the whole CU7 team. The photometric data reduction was done with the PhotPipe pipeline at DPCI; processing data were received from the IDT pipeline at DPCE.
EPSL data examples

RR-Lyrae stars in the Large Magellanic Cloud

Gaia Image of the week: http://www.cosmos.esa.int/web/gaia/iow_20150305

Gaia data releases

<table>
<thead>
<tr>
<th>Year</th>
<th>Release 1</th>
<th>Release 2</th>
<th>Release 3</th>
<th>Release 4</th>
<th>Final release</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>alpha and delta, mean G-magnitude</td>
<td>5-parameter astrometric solutions for single star (parallax)</td>
<td>Mean V_{rad}</td>
<td>Variable stars classification (All variability products)</td>
<td>everything!</td>
</tr>
<tr>
<td>2015</td>
<td>100K proper motion stars (Hipparcos+Gaia)</td>
<td>Integrated BP/RP + Astrophysical parameters</td>
<td>5-par astrometry</td>
<td>non-single star catalogue</td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td>Adequately characterised and calibration Ecliptic-pole data</td>
<td>Mean V_{rad} (for non variable)</td>
<td>Object classifications and Astrophysical Parameters</td>
<td>solar system objects</td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td></td>
<td></td>
<td>Orbital solution of binaries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td></td>
<td></td>
<td>mean RVS spectra</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2021</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2022</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Science operations started Today!

Nominal mission end extended mission end?

Release 1 and/or 2:
- Possibly: RR-Lyrae, Cepheids, ...

Release 1:
- alpha and delta, mean G-magnitude
- 100K proper motion stars (Hipparcos+Gaia)
- Adequately characterised and calibration Ecliptic-pole data

Release 2:
- 5-parameter astrometric solutions for single star (parallax)
- Integrated BP/RP + Astrophysical parameters
- Mean V_{rad} (for non variable)

Release 3:
- Mean V_{rad}
- 5-par astrometry
- Object classifications and Astrophysical Parameters
- Orbital solution of binaries
- mean RVS spectra

Release 4:
- Variable stars classification (All variability products)
- non-single star catalogue
- solar system objects

Final release:
- everything!

Berry Holl (Department of Astronomy, University of Geneva)