Variability Analysis and Processing of Gaia EPSL data

Laurent Eyer on behalf of Coordination Unit 7

Liverpool, UK

Tuesday November 11, 2015
The Gaia multi-epoch survey

Gaia is an exceptional survey, because:

- Gaia furnishes exquisite parallaxes
- Gaia surveys the entire sky with one set of instruments (G photometry, BP, RP spectrophotometry, RVS spectrometry)
- Gaia performs nearly simultaneous measurements in these different instruments
- Gaia measures the “bright sky”
- Gaia has a peculiar sampling

This talk is “only” about G magnitude
Two concepts: 1. Coordination Units (CU)
2. Data Processing Centers (DPC)
**Operation Rehearsals**

*Satellite has been in operation* (since July 2014)

Variability Processing and Analysis is *not* “in operation” *(yet)*

We have been *training*…. on

— “Gaia” Simulated data
— some real data (Hipparcos, OGLE, EROS)

And in the beginning of 2015 on

— **Real Gaia Ecliptic Pole Scanning Law data**

This exercise/training was called **Operation Rehearsal (OR5 stage 2)**

Official goal of the operation rehearsal: not to get science results
Goal is to turn the crank for the software

Variability Processing and Analysis: about **590,000 lines** of Java-R-SQL-XML-XSD code
The Operation Rehearsal data

Data set

- 28 days of Ecliptic Scanning law
- 3 days of Nominal Scanning law

69 million sources received from Photometric Processing (Cambridge University)

Selection:
- 20 measurements in either G, BP, RP
- No repeated observations within 100 minutes

790,000 sources
Mean number of obs/source (789K)

Map of number of unfiltered observations (equatorial coordinates, 1pix = 0.84 deg²)
Catalog GAIA-OR5S2-NO-REPEAT-GT20FOV, band
Gaia 789K + EROS2, OGLE, Hipparcos, Planetary Transits

(Equatorial coordinates, deg)

Courtesy of L. Rimoldini
+ Catalina (periodic)

(Equatorial coordinates, deg)

Courtesy of L. Rimoldini
+ MILLIQUAS

(Equatorial coordinates, deg)

Courtesy of L. Rimoldini
SEP: Gaia + OGLE

(Equatorial coordinates, deg)

Courtesy of L. Rimoldini
SEP: Gaia + OGLE, EROS2, Hipparcos, AAVSO, MILLIQUAS

(Equatorial coordinates, deg)
SEP: Gaia + OGLE, EROS2, Hipparcos, AAVSO, MILLIQUAS, Planetary Transits

Courtesy of L. Rimoldini
To get the data flavour
Comparison with OGLE

Image of the Week (March 05): RR Lyrae stars

CU7 / DPCG Variability Analysis

Calibrated photometry (CU5)
Radial velocities (CU6)

General Variability Detection (GVD) → Characterization
Special Variability Detection (SVD) → Characterization

Specific Object Studies (SOS) → Classification

Variables catalogue (CU7) → Global Variability Studies (GVS)

Unexpected Features Analyses
Supplementary Observations

Astrometric char (CU3+CU4)
Spectroscopic char (CU6)
Astrophysical param (CU8)
CU7 processing chain

OR5 Stage 2 EPSL (Dec '14-Mar '15)
Greyed out: not run this OR.

Calibrated photometry (CU5)

- Statistical parameters
- + p-values of CU5

General Variability Detection (GVD)
- p-value computations + classifier

Special Variability Detection (SVD)
- short-time
- solar-like
- planets

Period search + Modeling

Attribute calculation
- Unsupervised
  - EB
  - Cep/RRL
  - LPV
    - AGN
    - PMS
- Supervised

Extractors
- transient
- μ-lens
- EB
- Be
- μ-lens
- CV
- Rapid-phases

Global Variability Studies (GVS)
- Quality assessment
- Bias Estimation

Unexpected Feature Analyses

Supplementary Observations

Courtesy of B.Holl
Two fundamental quantities to estimate:

— **Completeness**
— **Contamination**

Detection was done with a classifier (Random Forest) attributes were computed a training set was defined (based on OGLE)

**Classifier result: The confusion matrix**

<table>
<thead>
<tr>
<th></th>
<th>VARIABLE</th>
<th>CONSTANT</th>
</tr>
</thead>
<tbody>
<tr>
<td>VARIABLE</td>
<td>376</td>
<td>80</td>
</tr>
<tr>
<td>CONSTANT</td>
<td>546</td>
<td>5</td>
</tr>
<tr>
<td>Contamination</td>
<td>8</td>
<td>13</td>
</tr>
</tbody>
</table>
Special Variability Detection:
Elisa Distefano, Shay Zucker, Brandon Tingley, Laurent Eyer, Isabelle Lecoeur, Maroussia Roelens, Leanne Guy, Alessandro Lanzafame
Special Variability Detection: short time scale

Laurent Eyer, Isabelle Lecoeur, Maroussia Roelens, Alessandro Lanzafame

Implementation of variogram: “variance” for all the paired magnitude difference separated by a certain time lag

One example of per-ccd data:
Special Variability Detection: exo-planet transits

Shay Zucker, Brandon Tingley, Leanne Guy, Alessandro Lanzafame

Two algorithms:

— Box-Least Square

— Outlier Probability, Tingley
  (A&A 2011)

The Box Least Square algorithm gives about 200 candidates

We do not claim any detection yet!

Conclusion Box-Least Square is functioning well
CU7 processing chain

OR5 Stage 2 EPSL (Dec '14-Mar '15)
Greyed out: not run this OR.

Calibrated photometry (CU5)

Statistical parameters

+ p-values of CU5

General Variability Detection (GVD)
- p-value computations + classifier

Special Variability Detection (SVD)
- short-time
- solar-like
- planets

Period search + Modeling

Attribute calculation

Unsupervised
- EB
- Cep/RRL
- LPV
- AGN
- PMS

Supervised

Extractors
- transient
- \(\mu\)-lens
- EB

Be
- \(\mu\)-lens

Flaring
- short-time

Rot. Mod.
- planets

Global Variability Studies (GVS)
- Quality assessment
- Bias Estimation

Unexpected Feature Analyses

Supplementary Observations

Rapid-phases

Statistical parameters

Courtesy of B. Holl
Characterisation
Jan Cuypers, Leanne Guy, Lorenzo Rimoldini, Joris De Ridder

Goal: To define attributes

- statistical parameters
- Modelling
  - Period search
  - Fourier Series and polynomial fit

Time series per object:

$\text{Time(i), G-, BP-, RP- \ mag(i) [ or RV/radial velocity(i) ] i=1,…, number of measurements}$
Few examples of modelling
Few examples of modelling
Calibrated photometry (CU5)

Statistical parameters

+ p-values of CU5

General Variability Detection (GVD)

p-value computations + classifier

Special Variability Detection (SVD)

short-time, solar-like, planets

Period search + Modeling

Attribute calculation

Unsupervised

Supervised

EB, Cep/RRL, LPV

EB, Cep/RRL, LPV

Unsupervised

Supervised

EB, Cep/RRL, LPV

Unsupervised

Supervised

EB, Cep/RRL, LPV

Unsupervised

Supervised

EB, Cep/RRL, LPV

Global Variability Studies (GVS)

Quality assessment, Bias Estimation

Period search + Modeling

Extractors

transient, μ-lens, EB

Be, μ-lens

CV, Rapid-phases

Supplementary Observations

OR5 Stage 2 EPSL (Dec '14-Mar '15)
Greyed out: not run this OR.

Statistics

Variability detection

Character

Classification

Specific Object Studies (SOS)

Additional Modules

Courtesy of B.Holl
Classification

Joris de Ridder, Berry Holl, Lorenzo Rimoldini, Luis Sarro, Sara Regibo, Mauro Lopez, Jonas Debosscher, Maria Sueveges

Supervised classification (several methods):

- Multistage tree: Bayesian networks
- Multistage tree: Gaussian mixture
- Random Forest

Tree for Gaussian Mixture:

- GDOR
- RRAB, RRC
- Cepheids: DCEP, DCEPS, CWA, CWB, CEP(B)
- Bcep
- DSCTC, DSCT+SXPHE

Furnish training set built from Crossmatched data
### Classification

Confusion matrix of Random Forest using cross-matched data (OGLE, Hipparcos, AAVSO, Milliquas)

<table>
<thead>
<tr>
<th></th>
<th>CONSTANT</th>
<th>QSO</th>
<th>ECL</th>
<th>OTHER</th>
<th>RRLYR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSTANT</td>
<td>79</td>
<td>8</td>
<td>5</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>QSO</td>
<td>7</td>
<td>88</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>ECL</td>
<td>4</td>
<td>3</td>
<td>84</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>OTHER</td>
<td>13</td>
<td>4</td>
<td>12</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>RRLYR</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>94</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>LPV</th>
<th>ELL</th>
<th>DSCT</th>
<th>DCEP</th>
<th>UG</th>
</tr>
</thead>
<tbody>
<tr>
<td>103</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>135</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>108</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>106</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Contamination | 34 | 20 | 19 | 34 | 5 | 43 | 100 | _ | _ | _
OR5 Stage 2 EPSL (Dec ‘14-Mar ‘15)
Number of sources (% of total).

Selection of CU5 photometry:
catalog: ‘GAIA OR5S2 NO-REPEAT GT20FOV’

789,472 per FOV  (785,664 also have per CCD)

Statistical parameters

General Variability Detection (GVD)
- p-value computations + classifier

52,719 (6.7%) short-time
- variogram: 19,210 (2.4%)

Special Variability Detection (SVD)
- solar-like
  - slope: 8,266 (1.0%)
  - variogram: 19,210 (2.4%)
- planets
  - 210 (0.03%)

Period search + Modeling

Found variable by GVD and/or SVD:

76,784 (9.7%)

Attribute calculation

Unsupervised
- EB
- Cep/RRL
- LPV

Supervised
- Flaring
- Rot. Mod.

Extractors
- transient
- μ-lens
- Be
- μ-lens

SOS

All: 3,465 (0.4%) Ceph: 0 (0 %) 139 (0.02%)
Selected: 729 (0.1%) RRL: 1,289 (0.2%)

Statistical parameters

SOS

139 (0.02%)
3,691 (0.5%)
210 (0.03%)
304 (0.04%)
215 (0.03%)
1,002 (0.1%)
Specific Object Studies
RR Lyrae and Cepheid stars
Gisella Clementini, Silvio Leccia, Vincenzo Ripepi, Nami Mowlavi, Isabelle Lecoeur

Image of the Week (May 28):

Classical overtone Cepheid
3 candidate anomalous Cepheids
Type 2 Cepheid

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.
Specific Object Studies: Eclipsing binaries

Eclipsing binaries go to CU4 for a full modelling
In CU7, some simple modelling are made
Solutions allow a ranking

Nami Mowlavi, Berry Holl, Isabelle Lecoeur, Fabio Barblan, Lorenzo Rimoldini

Highest rank
Specific Object Studies: Eclipsing binaries

Nami Mowlavi, Berry Holl, Isabelle Lecoeur, Fabio Barblan, Lorenzo Rimoldini

Lowest rank
One example: Comparison of distribution functions of RR Lyrae stars
Release scenario

spacecraft operations start

nominal mission end

extended mission end?


Release 1:
- $\alpha$ and $\delta$, mean G-magnitude
- Commissioning data
- 100K proper motion stars (Hipparcos +Gaia)

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
- Non-single star catalogue
- Solar system objects

Final release:
- everything!

Courtesy of B.Holl
Conclusions

- We can remark that for a first reduction, the quality of G band is remarkable.

- The photometry has improved already (for the current Operation Rehearsal).

- The Variability Processing and Analysis seems to be on the right track!

- We (CU5/CU7 joint effort) may release some EPSL data variability analysis (Data Release 1).