The challenges analysing Gaia Time Series

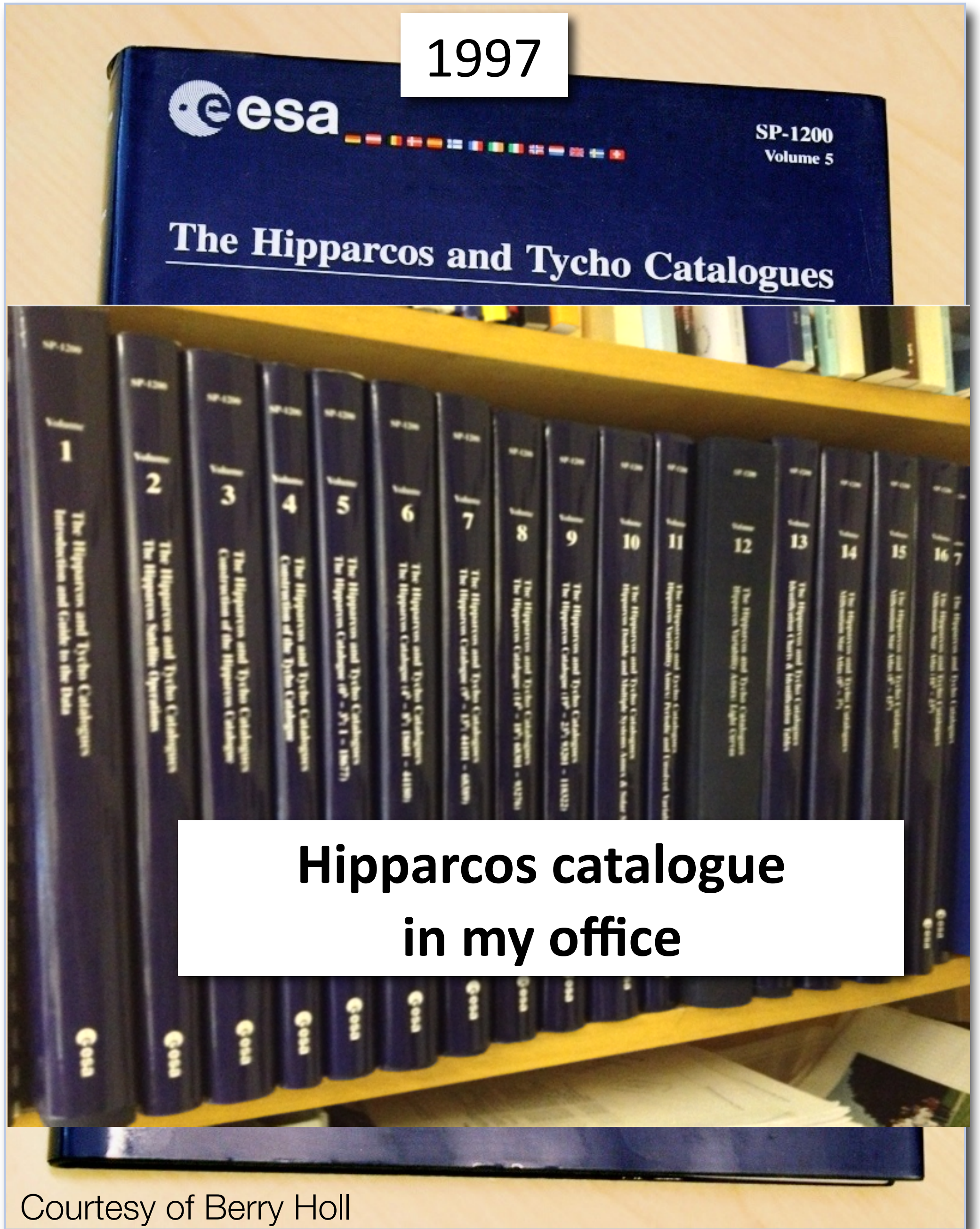
Laurent Eyer[1]

Nami Mowlavi[1], Dafydd W.Evans[2], Berry Holl[1], Lorenzo Rimoldini[1], Alessandro Lanzafame[3], Leanne Guy[1], Shay Zucker[4], Brandon Tingley[5], Isabelle Lecoeur-Taïbi[1], Maroussia Roelens[1], Jan Cuypers[6], Joris De Ridder[7], Sara Regibo[7], Manuel López[8], Jonas Debosscher[7], Maria Süveges[1], Luis Sarro[9], Gisella Clementini[10], Silvio Leccia[11], Vincenzo Ripepi[11], Fabio Barblan[1] , André Moitinho[12], Krzysztof Nienartowicz[1], Diego Ordoñez-Blanco[1], Jonathan Charnas[1], Grégory Jévardat de Fombelle[1]

[1]Department of Astronomy, University of Geneva, Versoix, Switzerland
[2]Institute of Astronomy, University of Cambridge, Cambridge, United Kingdom
[3]Dipartimento di Fisica e Astronomia, Universita di Catania, Catania, Italy
[4]Department of Geosciences, Tel Aviv University, Tel Aviv, Israel
[5]Institut for Fysik og Astronomi, Aarhus Universitet, Aarhus, Denmark
[6]Royal Observatory of Belgium, Brussels, Belgium
[7]Instituut voor Sterrenkunde, KU Leuven, Leuven, Belgium
[8] Centro de Astrobiologia, Departamento de Astrofisica, Villanueva de la Canada, Spain
[9]Departamento de Inteligencia Artificial, UNED, Madrid, SpainI
[10]INAF Osservatorio Astronomico di Bologna, Bologna, Italy
[11]INAF-Osservatorio Astronomico di Capodimonte, Napoli, Italy
[12]Faculdade de Ciencias de Universidade de Lisboa, Lisboa, Portugal

Statistical Challenges in Modern Astronomy
Carnegie Mellon University, Pittsburgh, USA
Wednesday June 8, 2016

Introduction: Ptolemy vs Hipparcos vs Gaia catalogues



1997

Hipparcos catalogue
in my office

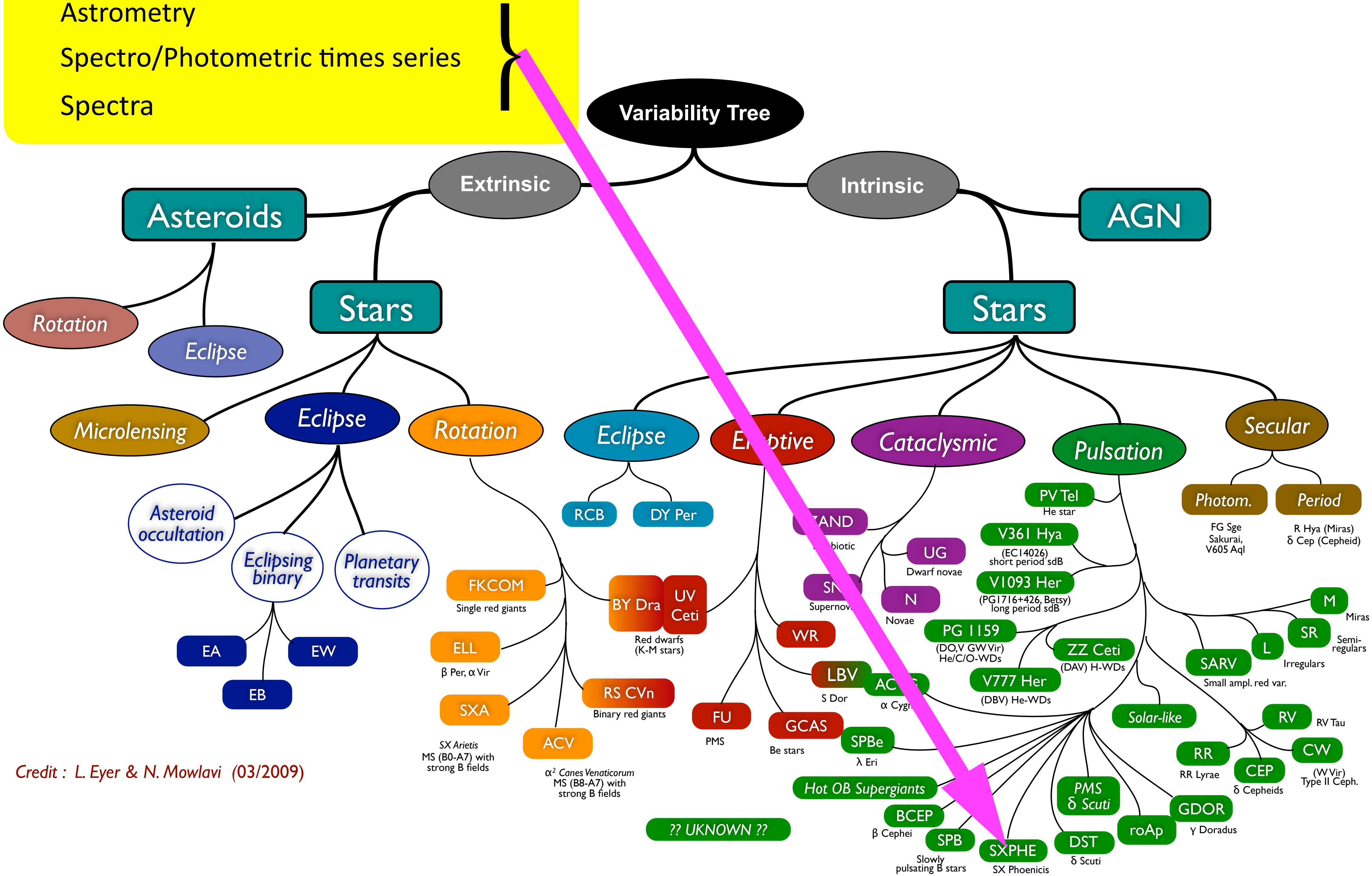
Courtesy of Berry Holl

Almagest
Ptolemy
~150 AD
Here ed. of
1528

Goal of Variability Processing and Analysis

Gaia measurements for source "i":

- Astrometry
- Spectro/Photometric times series
- Spectra

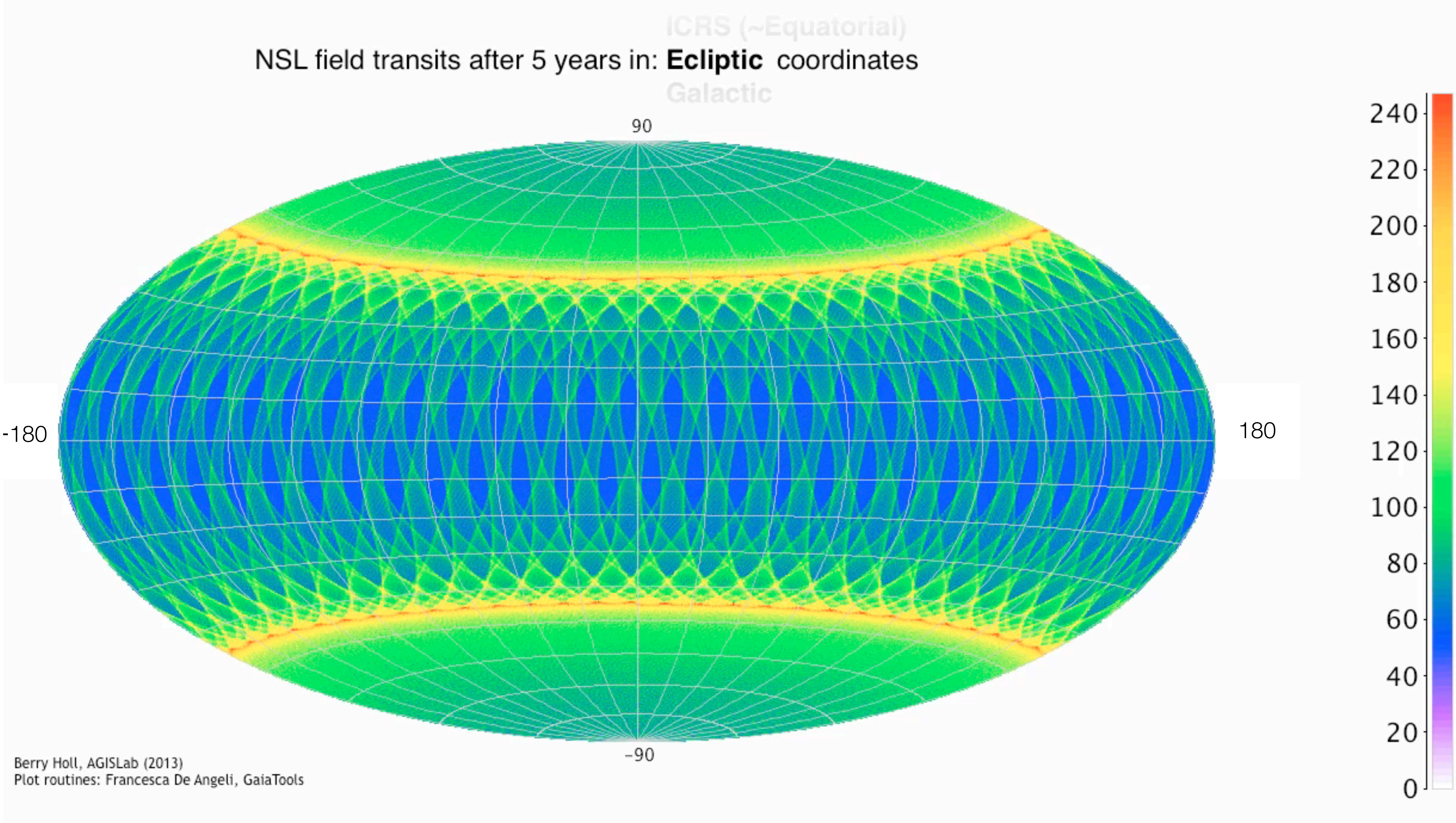


Credit : L. Eyer & N. Mowlavi (03/2009)

Challenges of Gaia in terms of time series analysis

- How to deal with large data set? (e.g. reshuffle data to get time series per source)
- How to gather qualitatively different quantities? (astrometry, photometry, spectra)
- How to handle different number of measurements (from 40 to 250), sparse and different samplings?
- How to deal with heteroscedastic data?
- How to do when there are poor estimates of uncertainties?
- How to search for “small” signals in ”noisy" data?
- How to perform model selection?
- How to compute the significance of peaks in a periodogram
- How to deal with the aliasing problem
- How to classify variable objects?
- How do we crossmath?
- How to rank objects?
- etc…

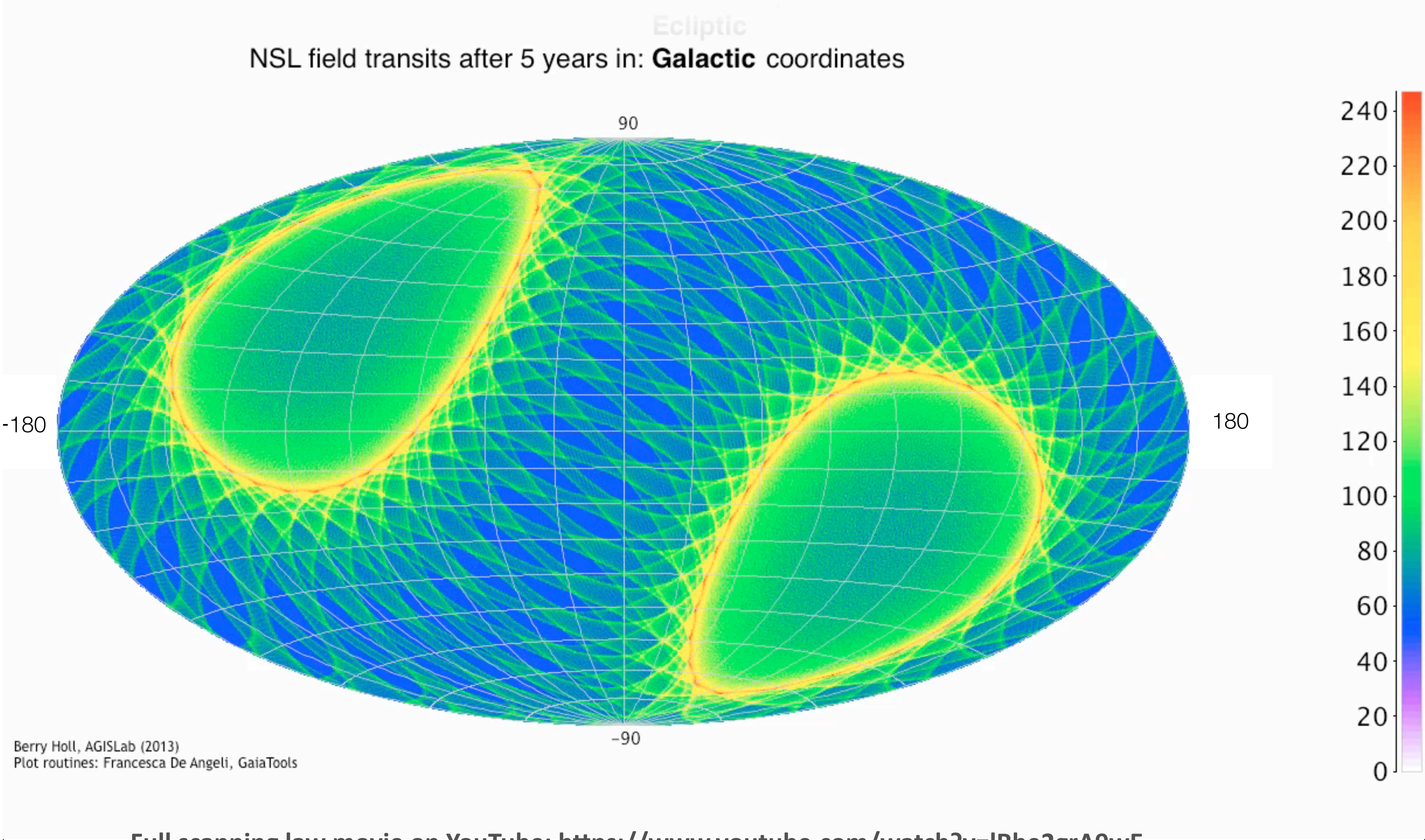
Five year nominal scanning law (NSL)



Full scanning law movie on YouTube: https://www.youtube.com/watch?v=lRhe2grA9wE

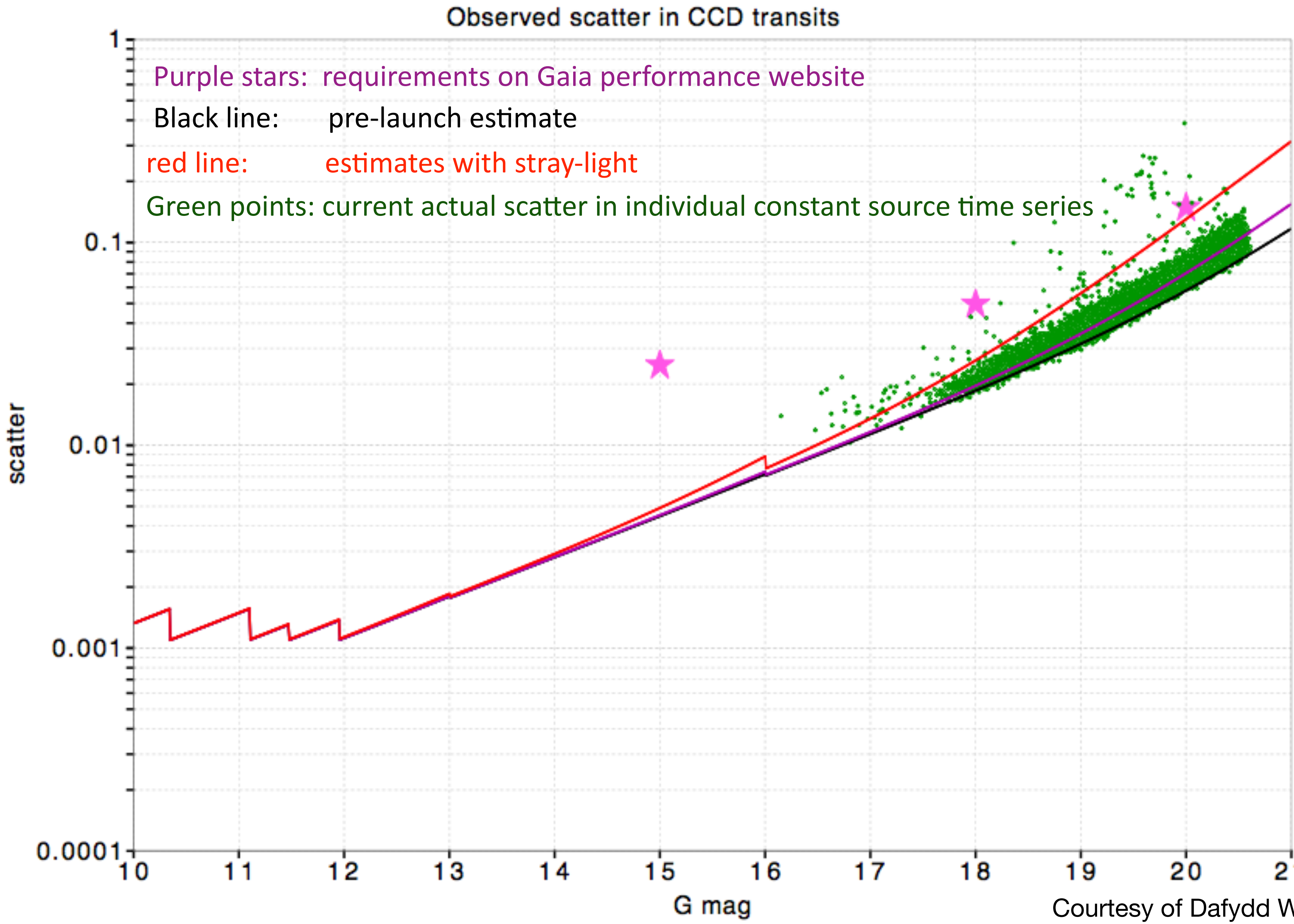
Courtesy of Berry Holl

Five year nominal scanning law (NSL)



Full scanning law movie on YouTube: https://www.youtube.com/watch?v=lRhe2grA9wE

Epoch photometric precision



Courtesy of Dafydd W.Evans

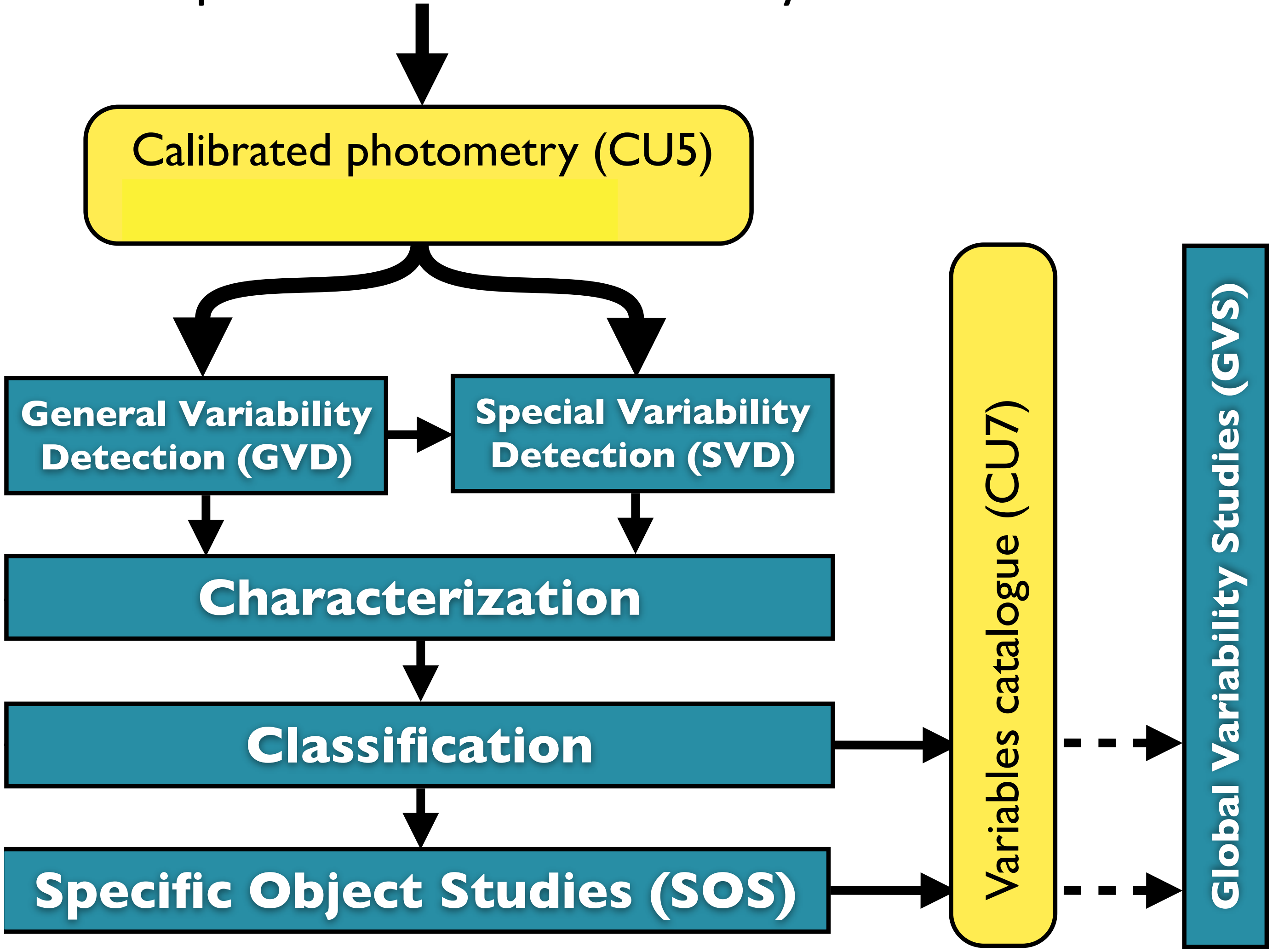
Pragmatic approach

- Develop a “method”
  - Test on simulated data (often very simple)
  - Test on real data
- Apply it to real Gaia data
  - Start on “easy“ signal, e.g. large signal to noise ratio
  - Start with law numbers of objects
- Interact with the data (subsamples) at all levels of the analysis

Gaia Variability Processing and Analysis

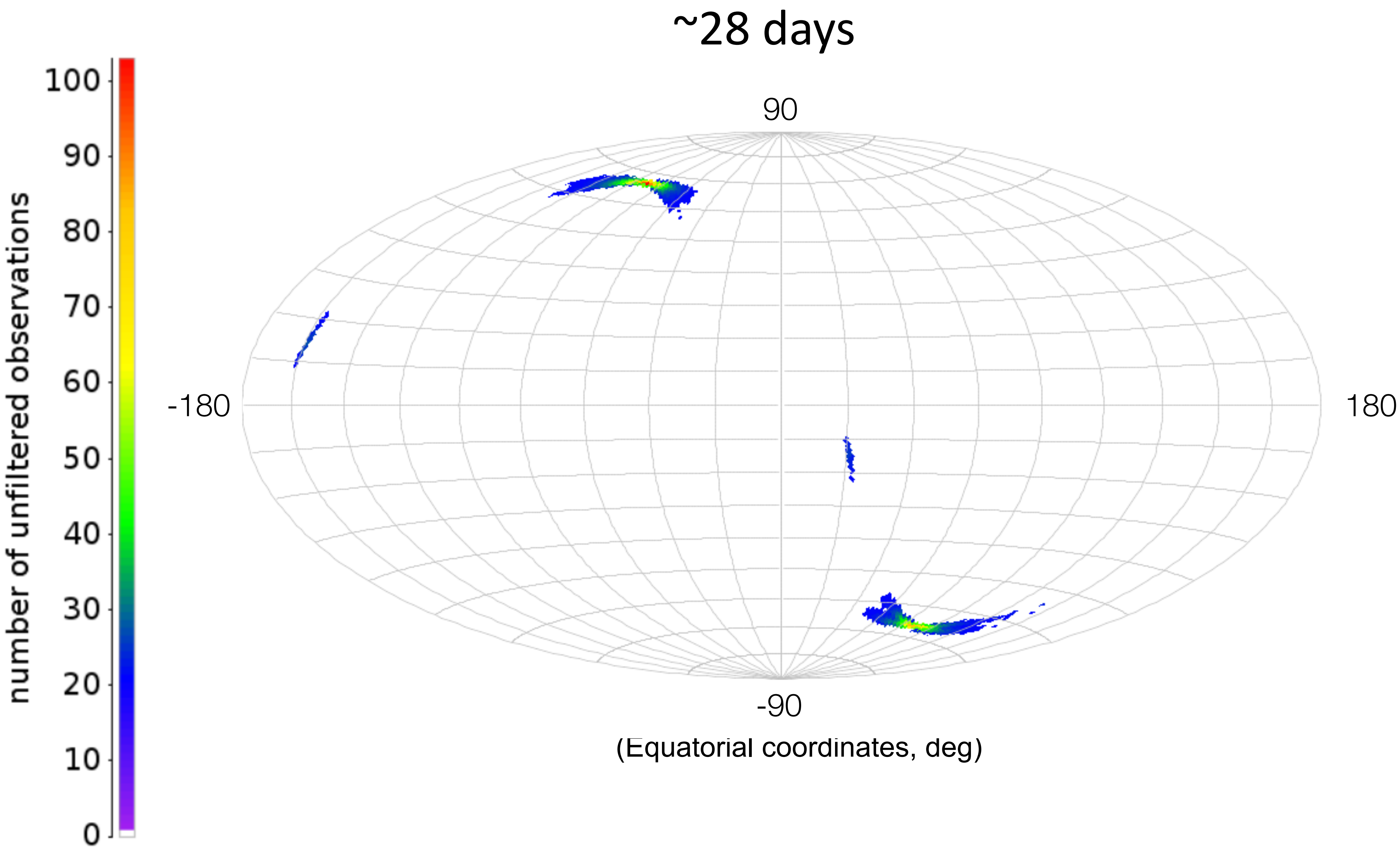
1 billion sources observed by Gaia
790,000 REAL Gaia data from Ecliptic Poles

Time series of 70 (40-250) measurements over 5 years
Time series up to 170 measurements over 28 days



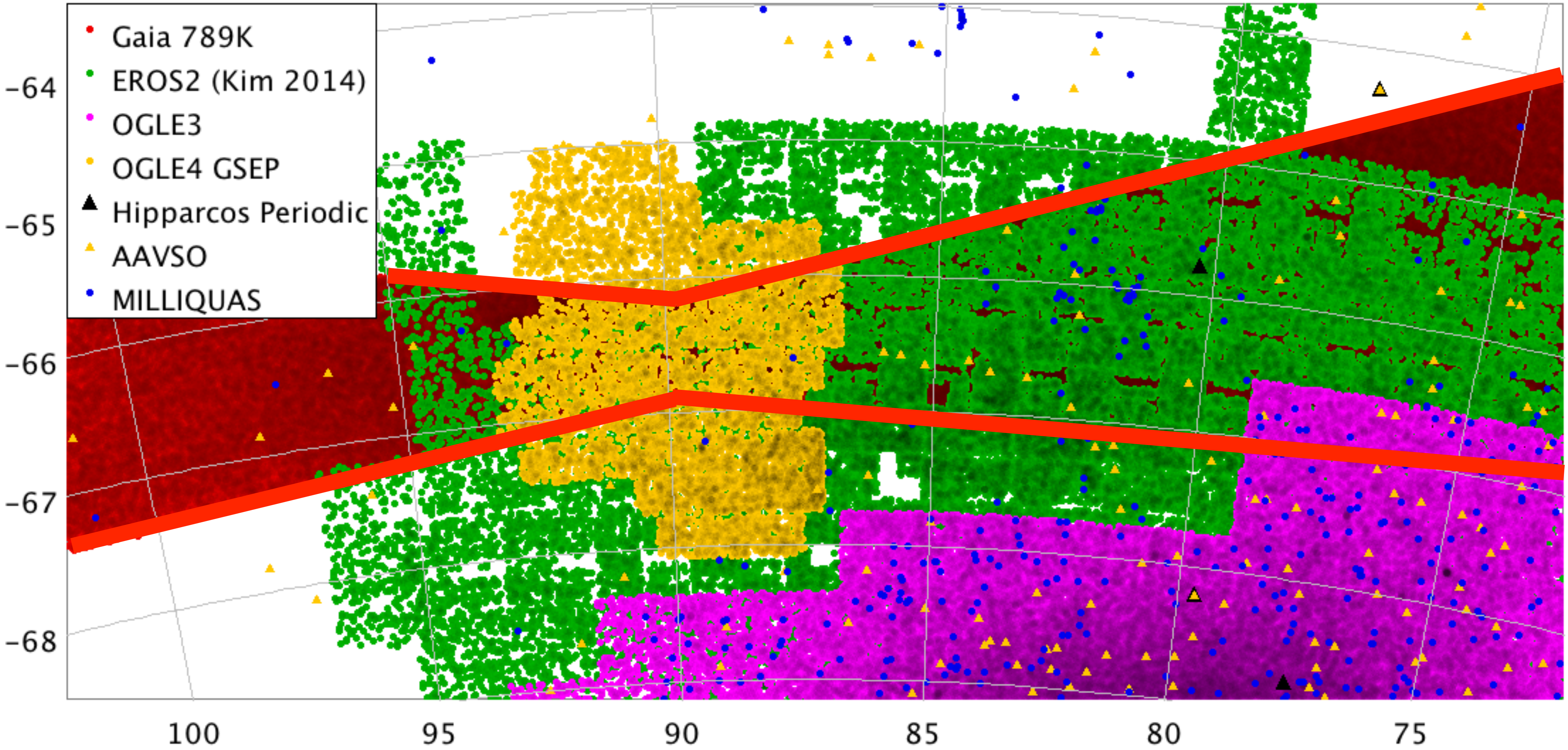
Unexpected Features Analyses

Selection of sources observed by Gaia from Ecliptic Pole Scanning Law (790,000 sources)



Warning: Sampling regular, different from Nominal Scanning law

South Ecliptic Pole region (part of Large Magellanic Cloud): Gaia and other surveys

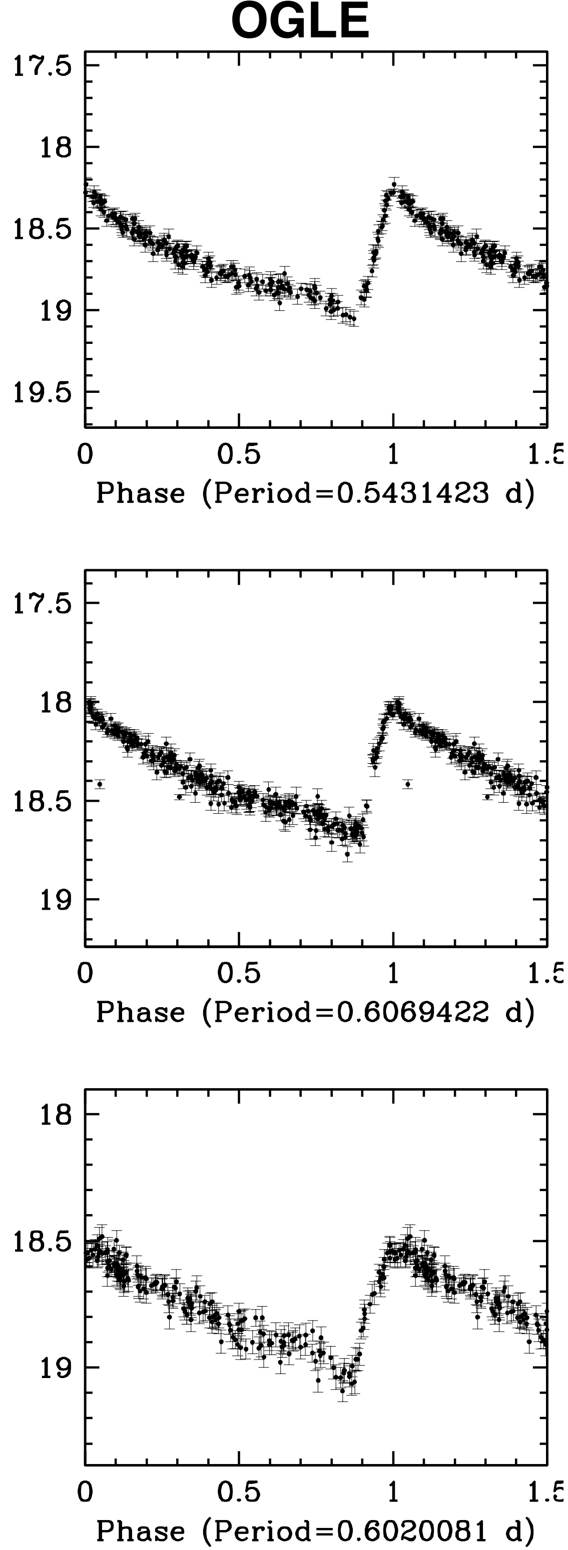


(Equatorial coordinates, deg)

To get the data flavour
Comparison with OGLE

Image of the Week (March 05, 2015):
RR Lyrae stars

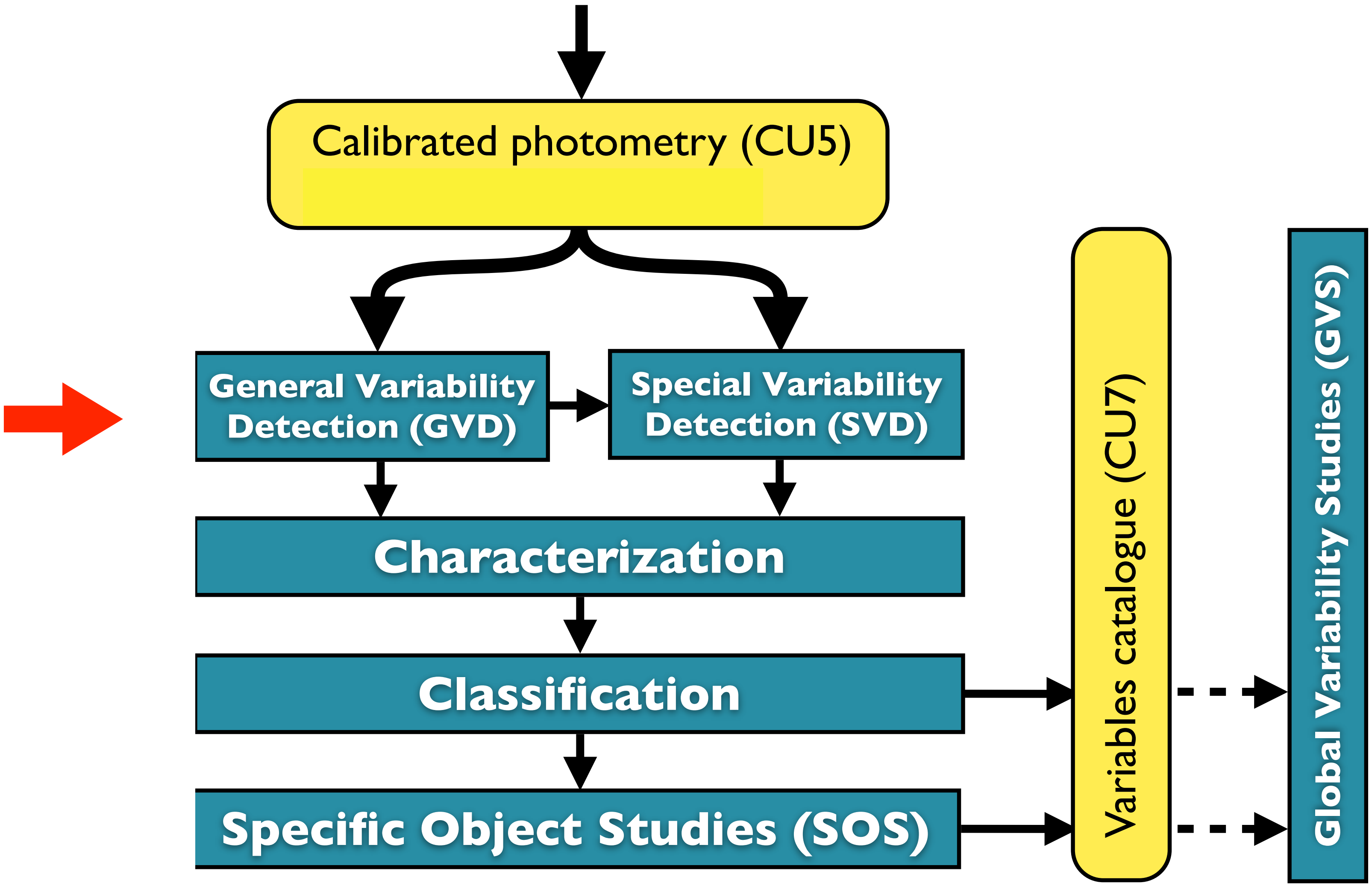
Credits: ESA/Gaia/DPAC/CU5/CU7/INAF-OABo, Gisella Clementini, Dafydd Evans, Laurent Eyer, Krzysztof Nienartowicz, Lorenzo Rimoldini and the Geneva CU7/DPCG and CU7/INAF-OACN teams.



Gaia Variability Processing and Analysis

790,000 REAL Gaia data from Ecliptic Poles

Time series up to 170 measurements over 28 days



General Variability Detection

Classical hypothesis testing can't be applied because of "poor" estimates of uncertainties

Detection was done with a classifier (Random Forest)

attributes were computed

a training set was defined (based on OGLE)

Two fundamental quantities to estimate:

—Completeness

—Contamination

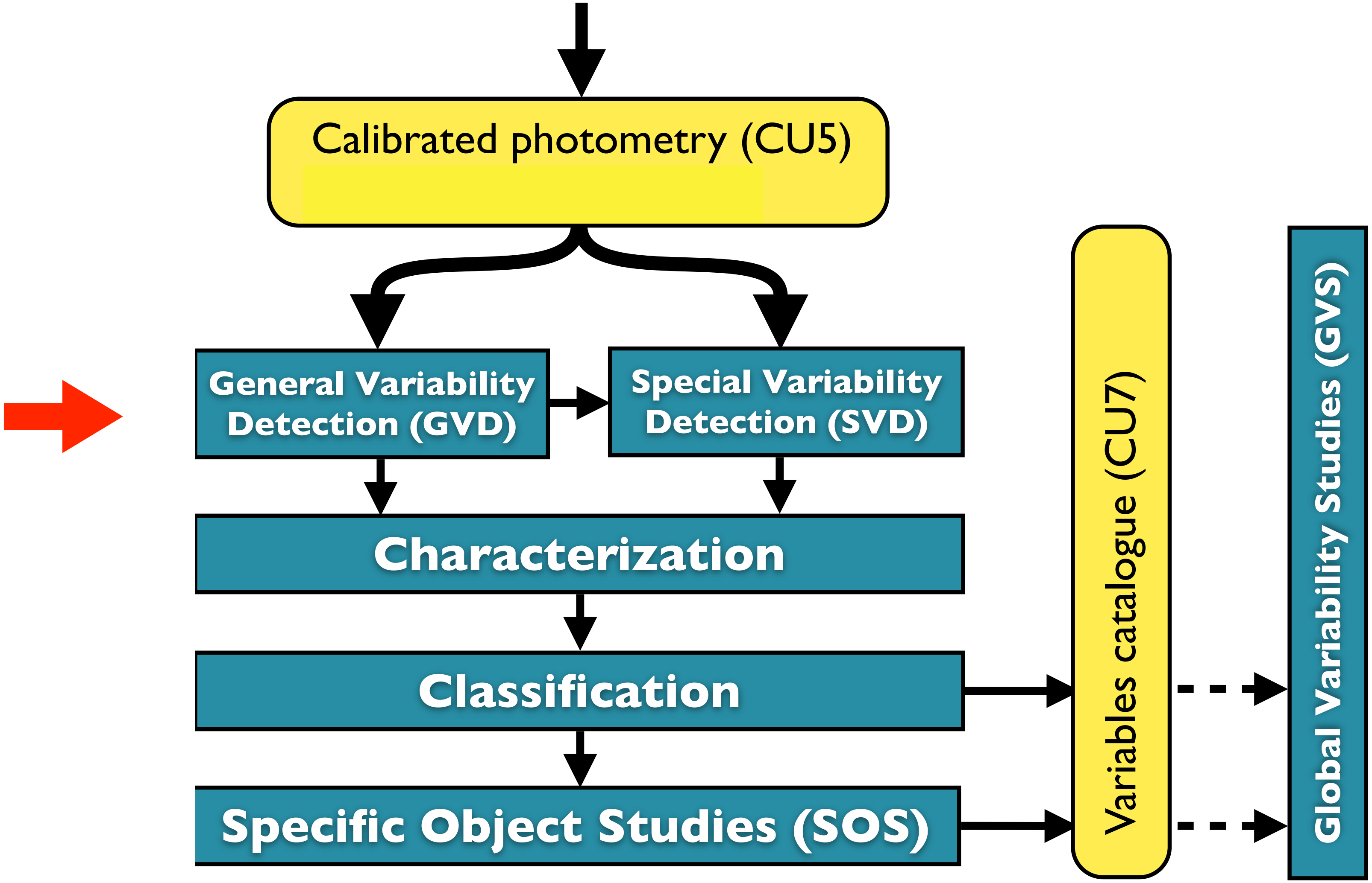
Classifier result: The confusion matrix

		VARIABLE	CONSTANT
376	VARIABLE	80	20
546	CONSTANT	5	95
	Contamination	8	13

Gaia Variability Processing and Analysis

790,000 REAL Gaia data from Ecliptic Poles

Time series up to 170 measurements over 28 days



Characterisation

Time series per object:

$\text{Time}_{(i)}$, G-, BP-, RP- $\text{mag}_{(i)}$ [or radial $\text{velocity}_{(i)}$]

i=1,…, number of measurements

Goal: To define attributes

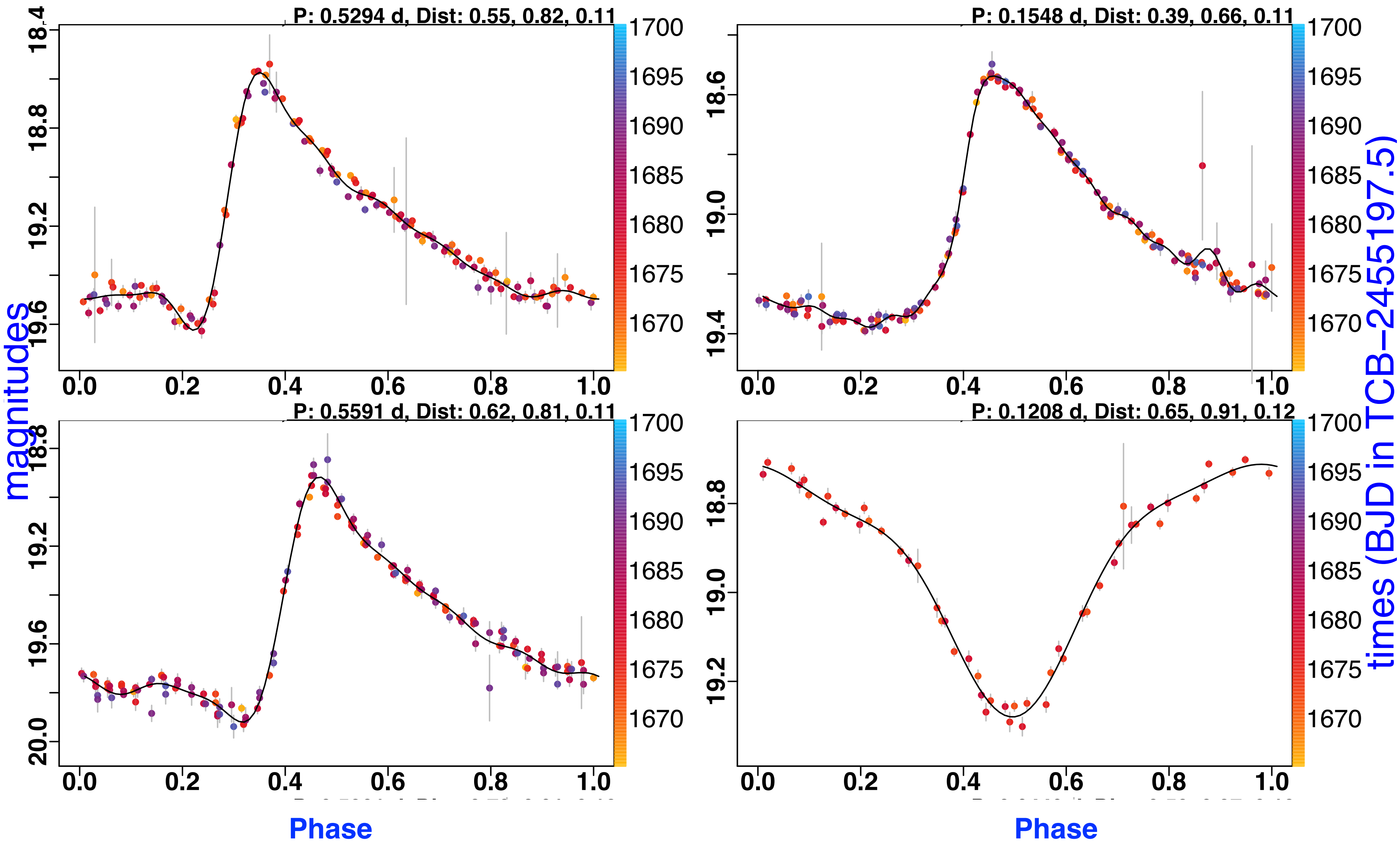
- statistical parameters

- Modelling

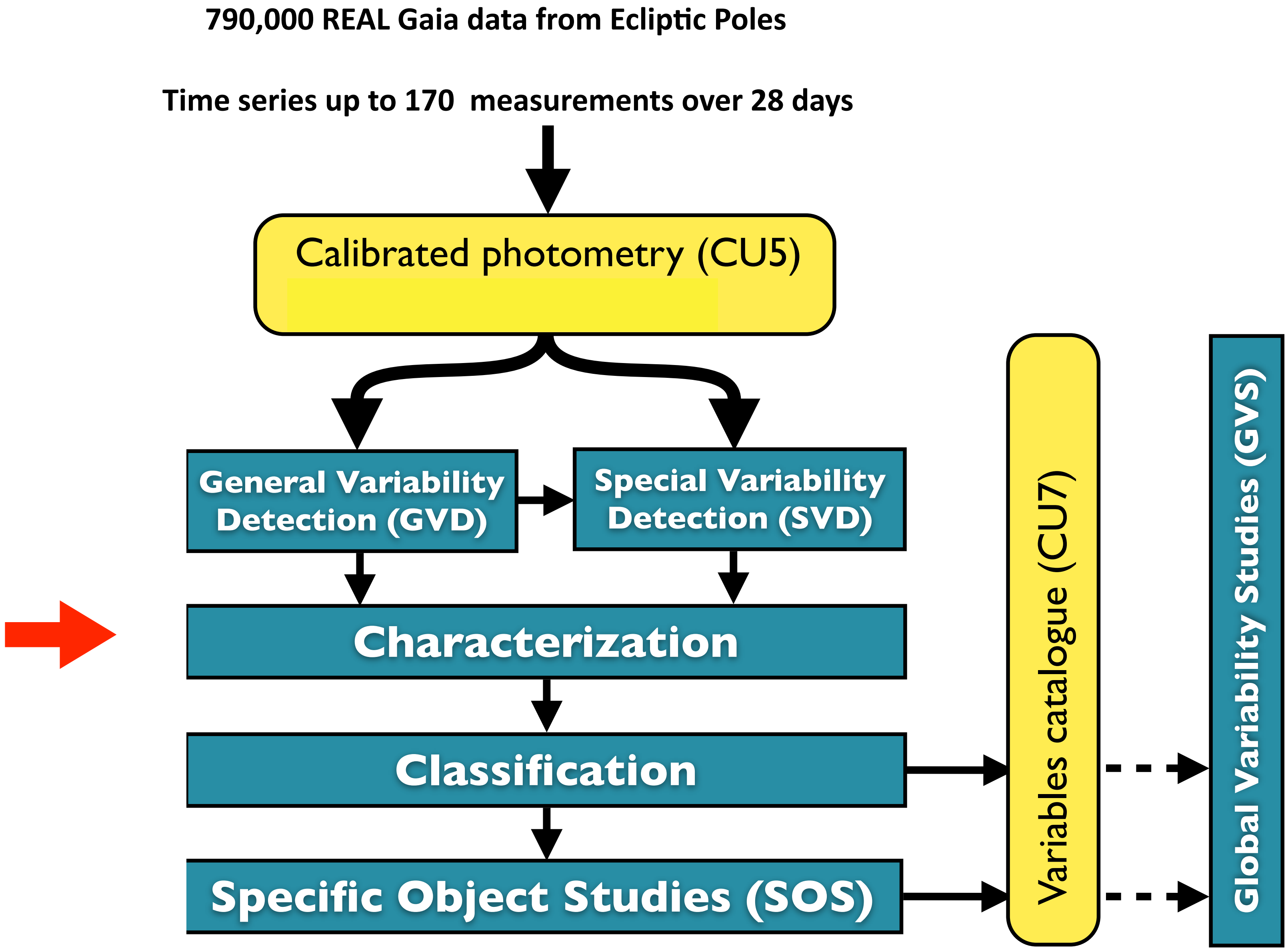
  —Period search

  —Fourier Series and polynomial fit

Characterisation: few examples of modelling



Gaia Variability Processing and Analysis



Classification

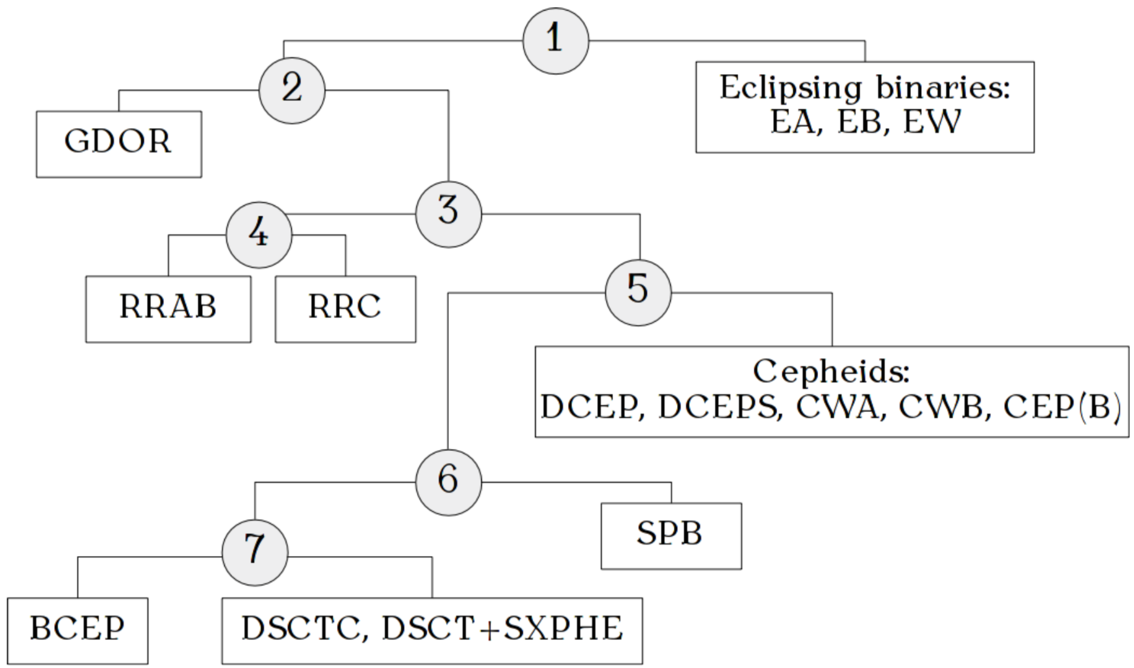
Supervised classification (several methods):

Multistage tree:
Bayesian networks

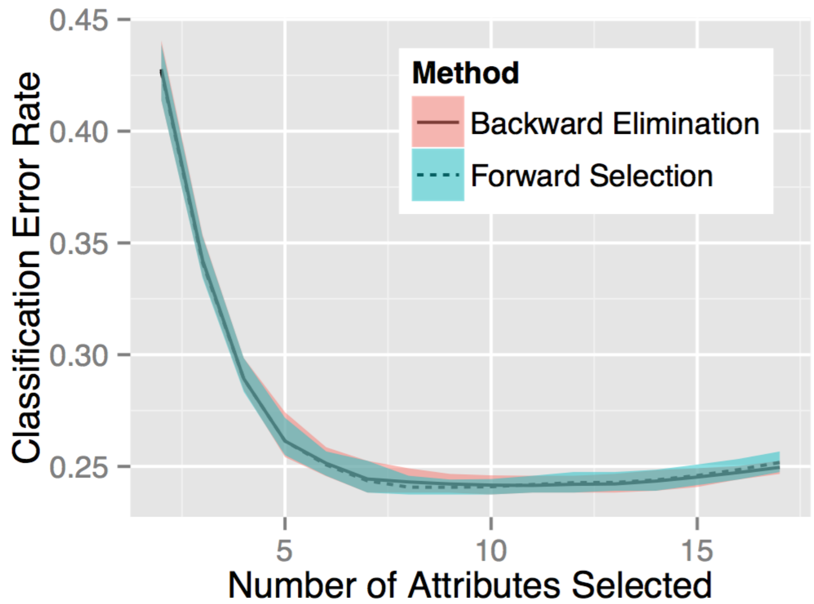
Multistage tree:
Gaussian mixture

Random Forest

Tree for Gaussian Mixture:



Furnish training set
built from Crossmatched data

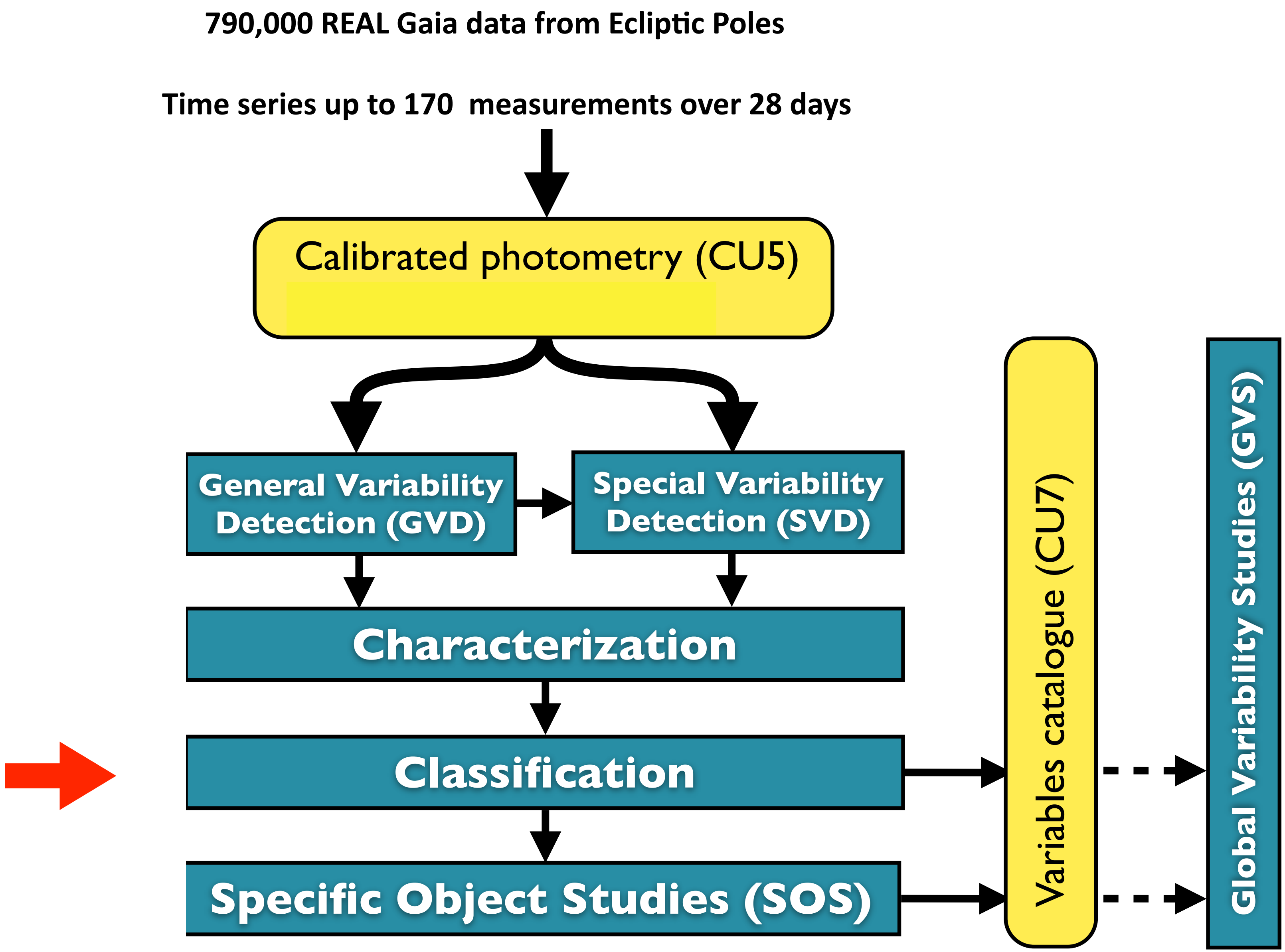


Classification

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

		CONSTANT	QSO	ECL	OTHER	RRLYR	LPV	ELL	DSCT	DCEP	UG
103	CONSTANT	79	8	5	9						
100	QSO	7	88	2	3						
135	ECL	4	3	84	7	1	1	1			
108	OTHER	13	4	12	69		2	1			
106	RRLYR	3		1	2	94					
27	LPV	30	19	7	30		15				
8	ELL	25	12		62						
7	DSCT	14		29	29	29					
2	DCEP	50		50							
2	UG					100					
	Contamination	34	20	19	34	5	43	100	–	–	–

Gaia Variability Processing and Analysis



Classification of RR Lyrae and Cepheid stars

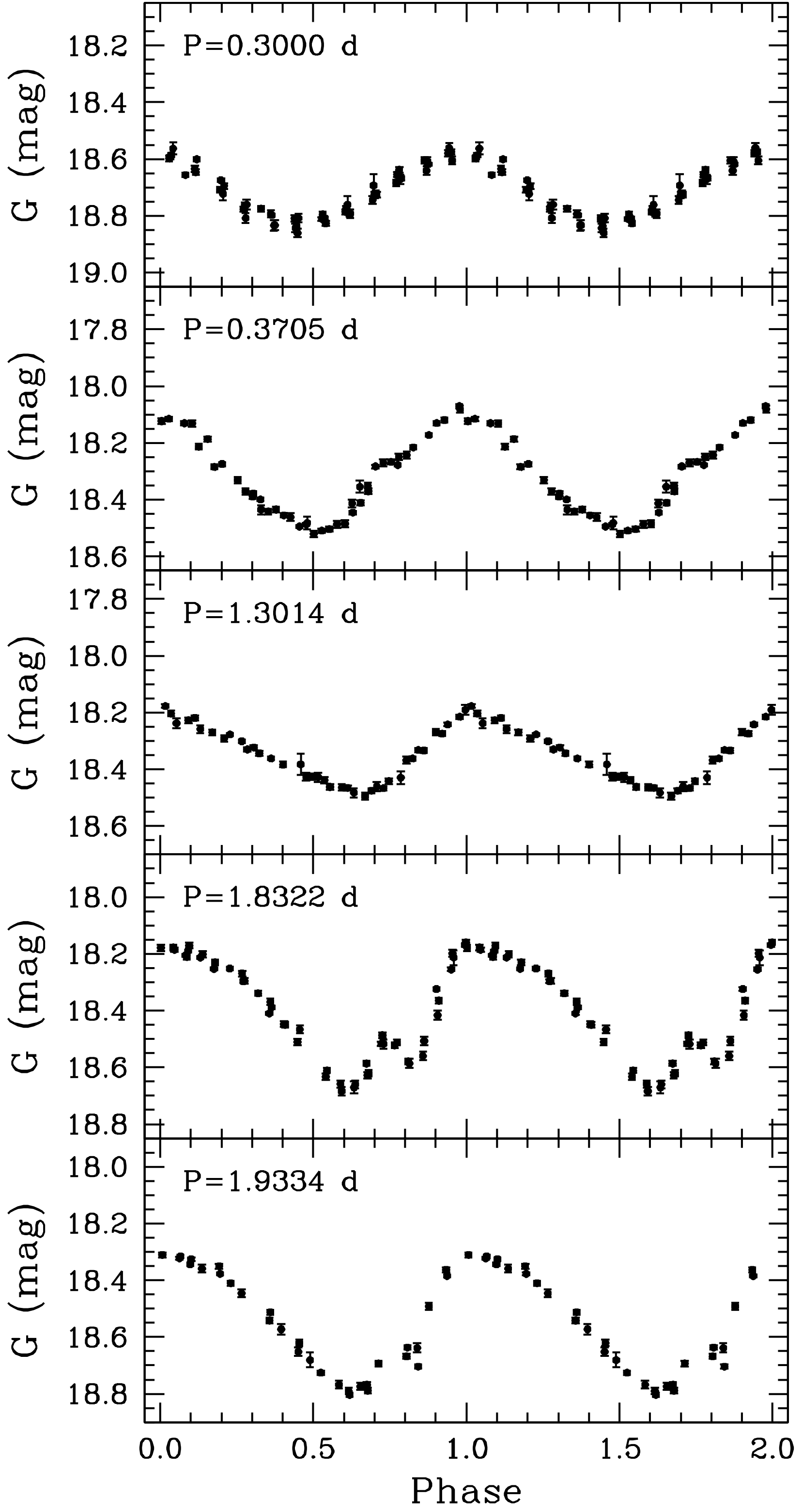
Gisella Clementini, Silvio Leccia, Vincenzo Ripepi, Nami Mowlavi, Isabelle Lecoeur

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.

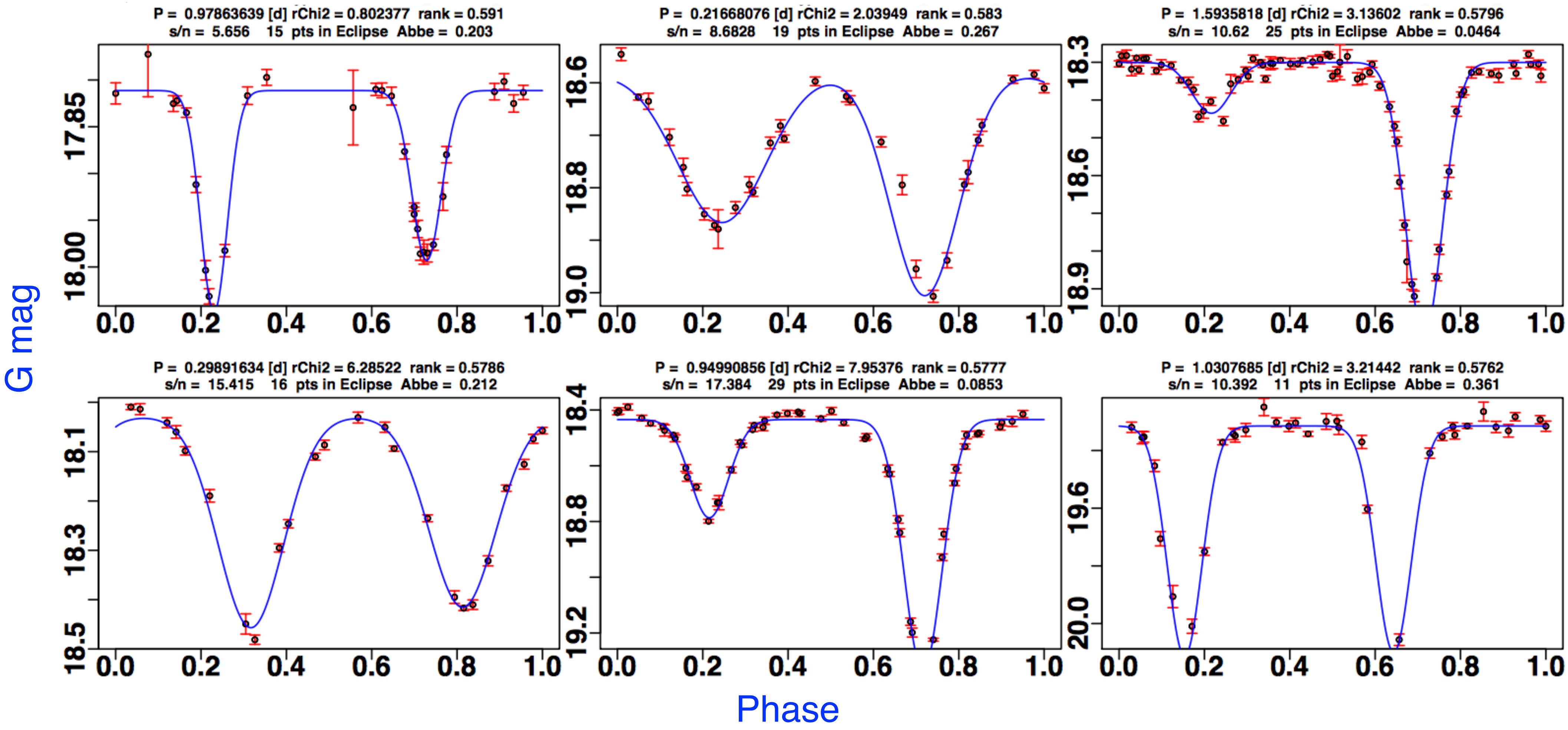


Specific Object Studies: Eclipsing binaries

Eclipsing binaries go to a dedicate treatment (Université Libre de Bruxelles) for a full modelling

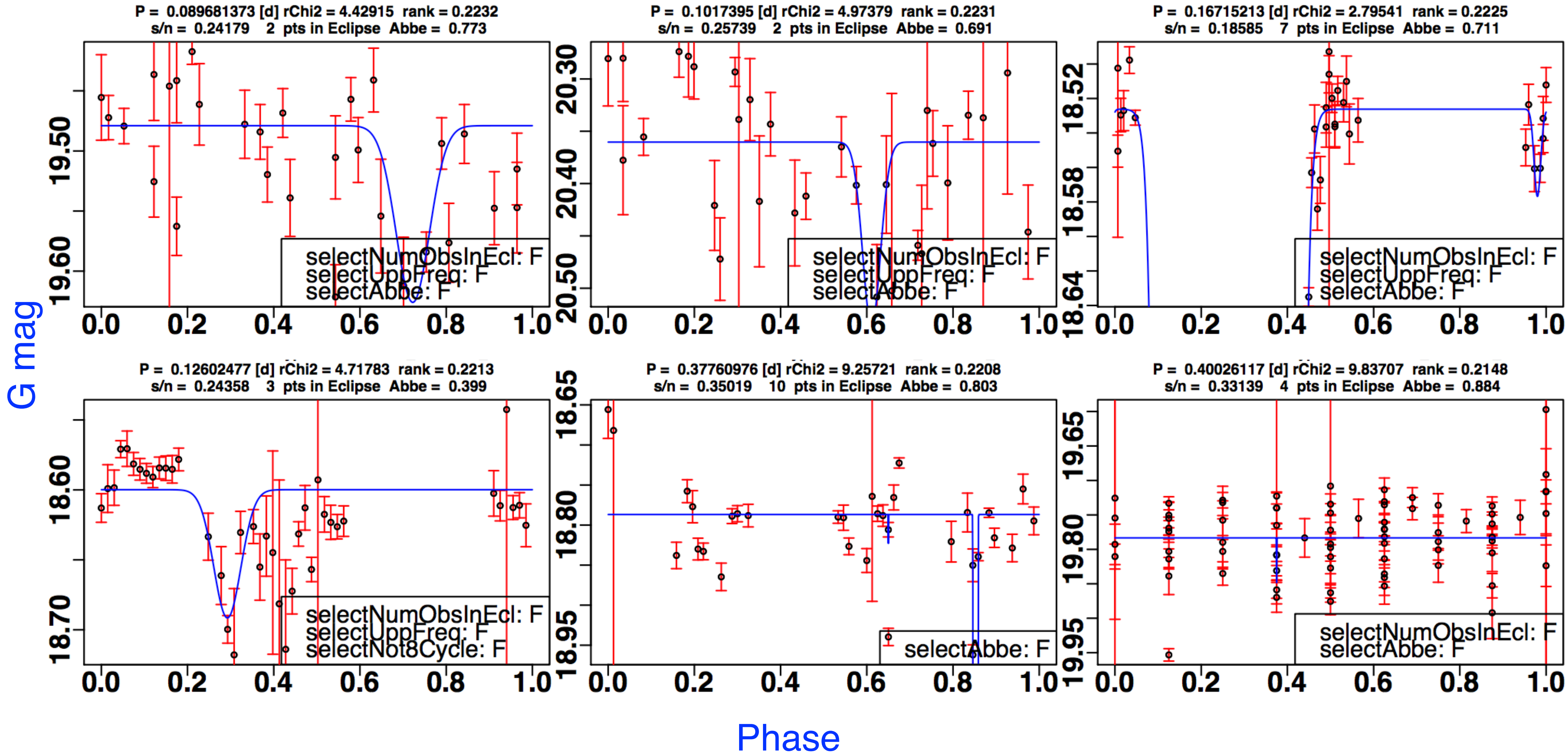
Here, simple modelling (fit with “two Gaussians”) are made
The solutions enable a ranking

Highest rank



Specific Object Studies: Eclipsing binaries

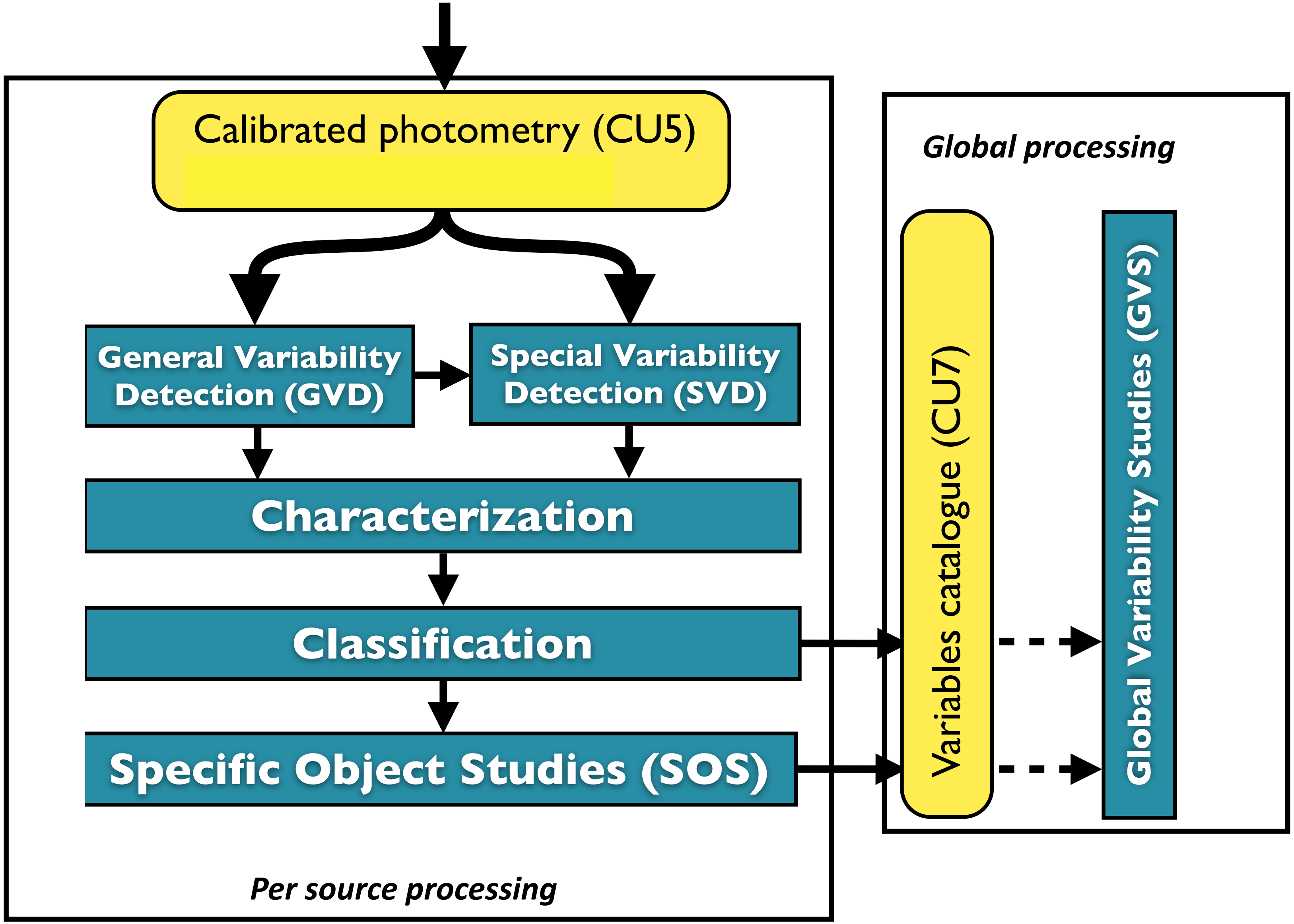
Lowest rank



Gaia Variability Processing and Analysis

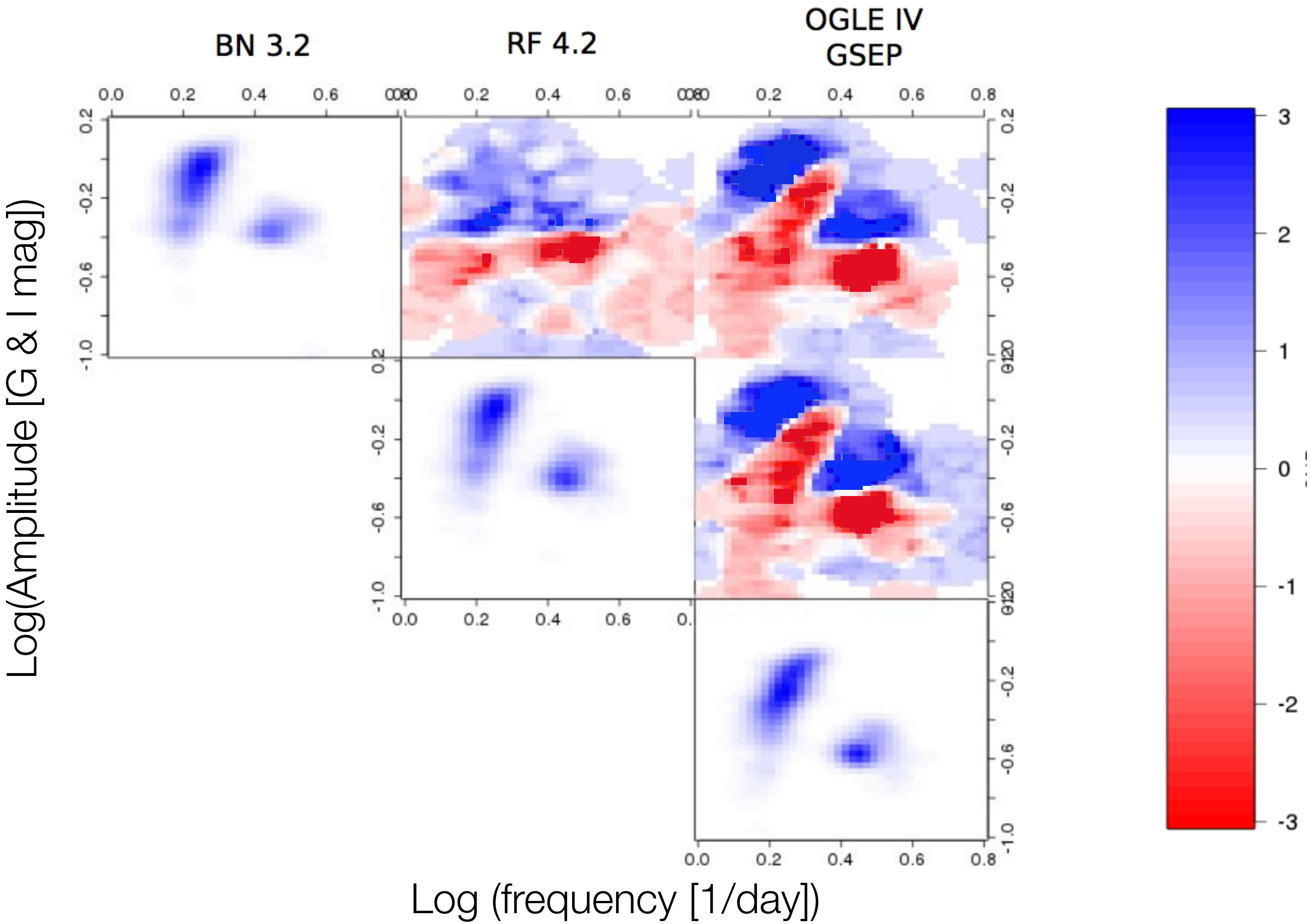
790,000 REAL Gaia data from Ecliptic Poles

Time series up to 170 measurements over 28 days



Global Variability studies

Comparison of distribution functions of RR Lyrae stars



Courtesy of L. Sarro (modified by L.Eyer)

The data releases

spacecraft operations start

nominal mission end

extended mission end?

2014	2015	2016	2017	2018	2019	2020	2021	2022

Release 1:

- **α and δ, mean G-magnitude for 1billion stars**
- ~~**100K proper motion stars (Hipparcos+Gaia)**~~

- **+ TGAS: Tycho-Gaia astrometric solution (2 million stars)**
- **+ Groups of variables: RR Lyrae stars & Cepheids (~3,300 stars)**

Release 2:

- **5-parameter astrometric solutions for single star (<u>parallax</u>)**
- **Integrated BP/RP + Astrophysical parameters**
- **Mean V_{rad} (for non variable)**

- **+ All sky: RR Lyrae stars & Cepheids & Short Time scale variables + ? (50,000-100,000 stars?)**

Release 3:

- Mean V_{rad}
- Object classifications and Astrophysical Parameters
- **Mean RVS, BP, RP spectra**
- **Orbital solution of binaries**

Release 4:

- **Variable stars classification**
- **Non-single star catalogue**
- **Solar system objects**

Final release:

- everything !

Courtesy of B.Holl

Data release caveats/limitations

- All sources have been treated as single stars
- Some filtering has been applied, e.g.;
  -Omit sources with too few observations, without astrometry and/or photometry, with very elongated positional error ellipses, …
  -Upper limit on errors in parallax, position, and photometry
- No high-proper-motion stars ($\mu > 3.5$ arcsec/year)
- Various unmodeled effects left in the data (chromaticity, CTI, micro-meteoroid hits, micro-clanks, …)
- Basic-angle-variation correction derived from on-board metrology
- Cross-matching limited by crude attitude, IGSL, and spurious sources
- Cyclic processing loops not yet closed
- All of these issues will be addressed in upcoming releases!

  Warning: the above weaknesses will lead to spatially-correlated systematics in DR1: do not blindly average astrometric quantities

Courtesy of Jos de Bruijne