

# Variability Analysis and Processing of Gaia EPSL data

Laurent Eyer on behalf of Coordination Unit 7

Tenerife, Spain

Tuesday June 23, 2015



**UNIVERSITÉ  
DE GENÈVE**



# Gaia Data Processing and Analysis Consortium

## DPAC

---

# Gaia Data Processing and Analysis Consortium

## DPAC

---

> 400 people

# Gaia Data Processing and Analysis Consortium

## DPAC

---

Two concepts:

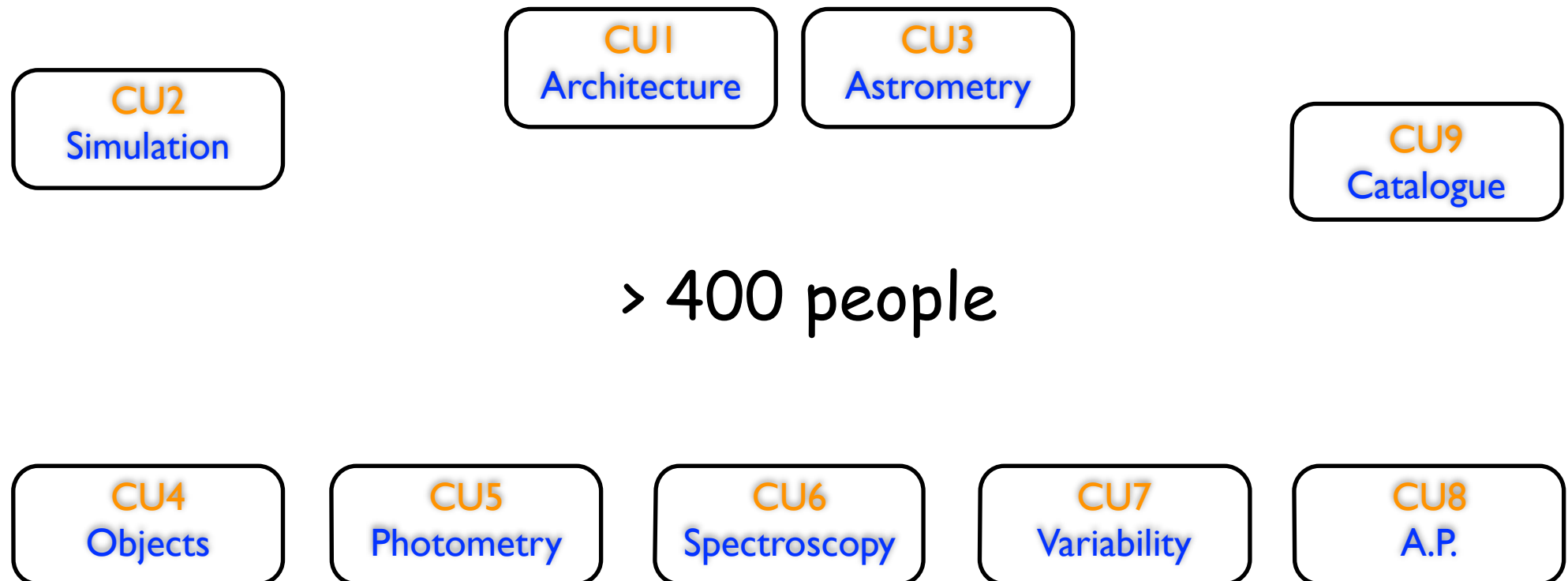
> 400 people



# Gaia Data Processing and Analysis Consortium DPAC

---

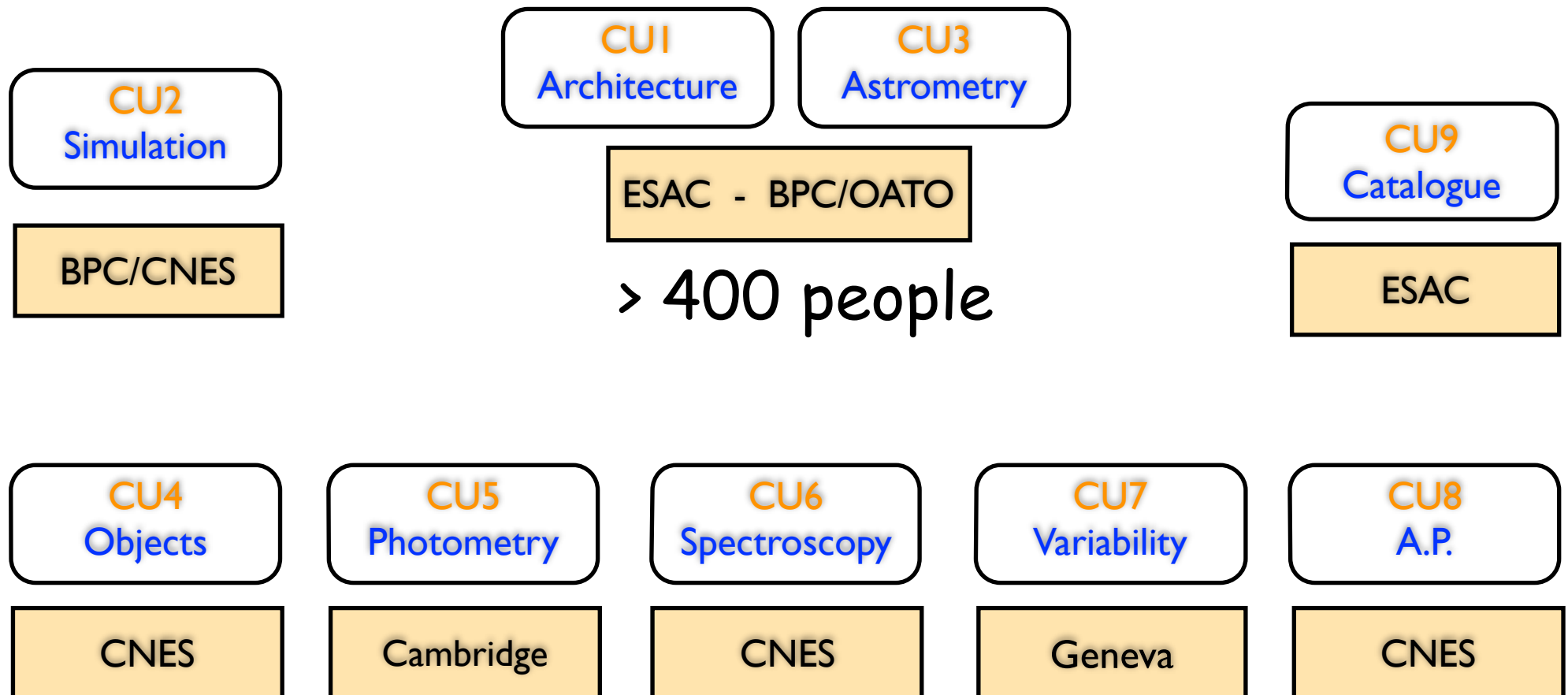
Two concepts: 1. Coordination Units (CU)



# Gaia Data Processing and Analysis Consortium DPAC

---

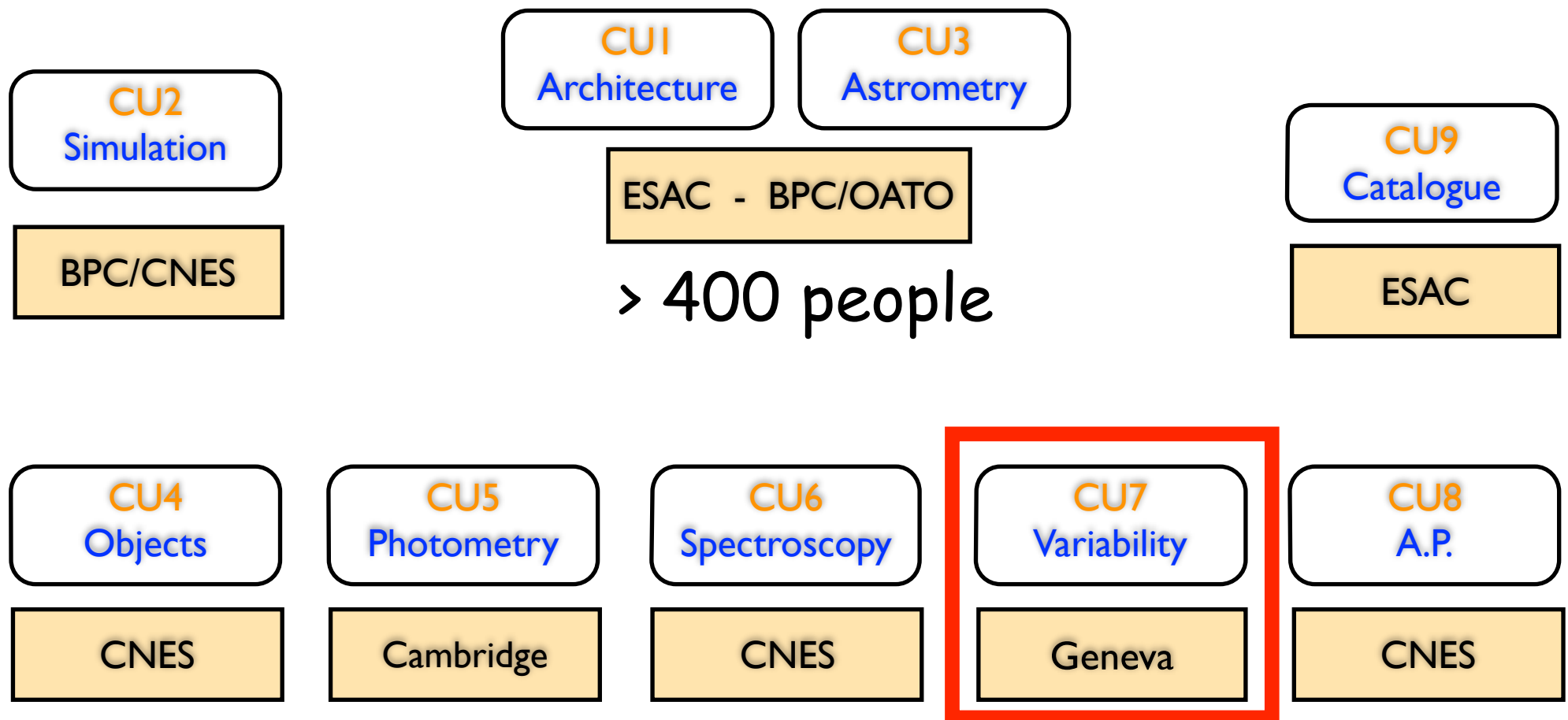
Two concepts: 1. Coordination Units (CU)  
2. Data Processing Centers (DPC)



# Gaia Data Processing and Analysis Consortium DPAC

---

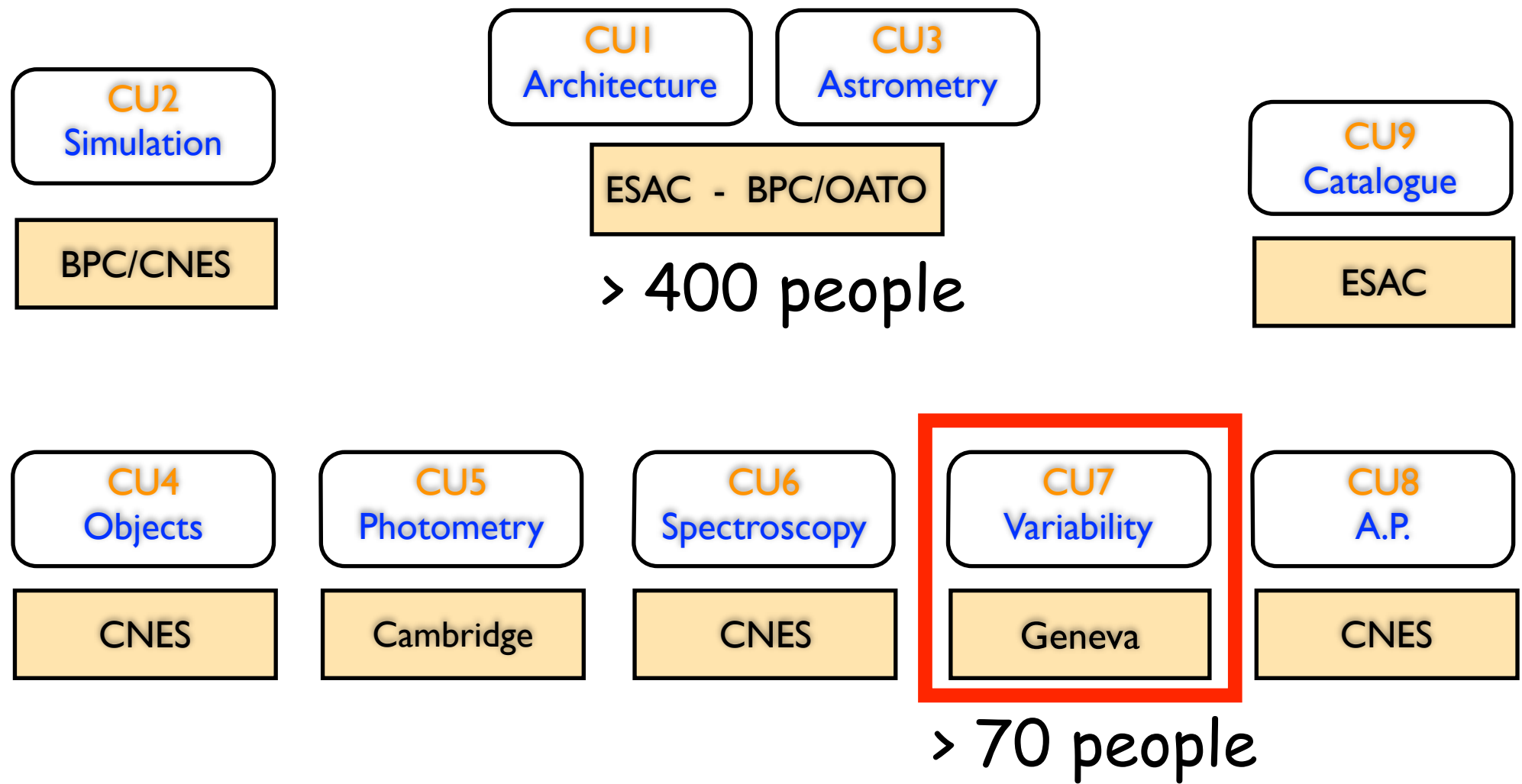
Two concepts: 1. Coordination Units (CU)  
2. Data Processing Centers (DPC)



# Gaia Data Processing and Analysis Consortium DPAC

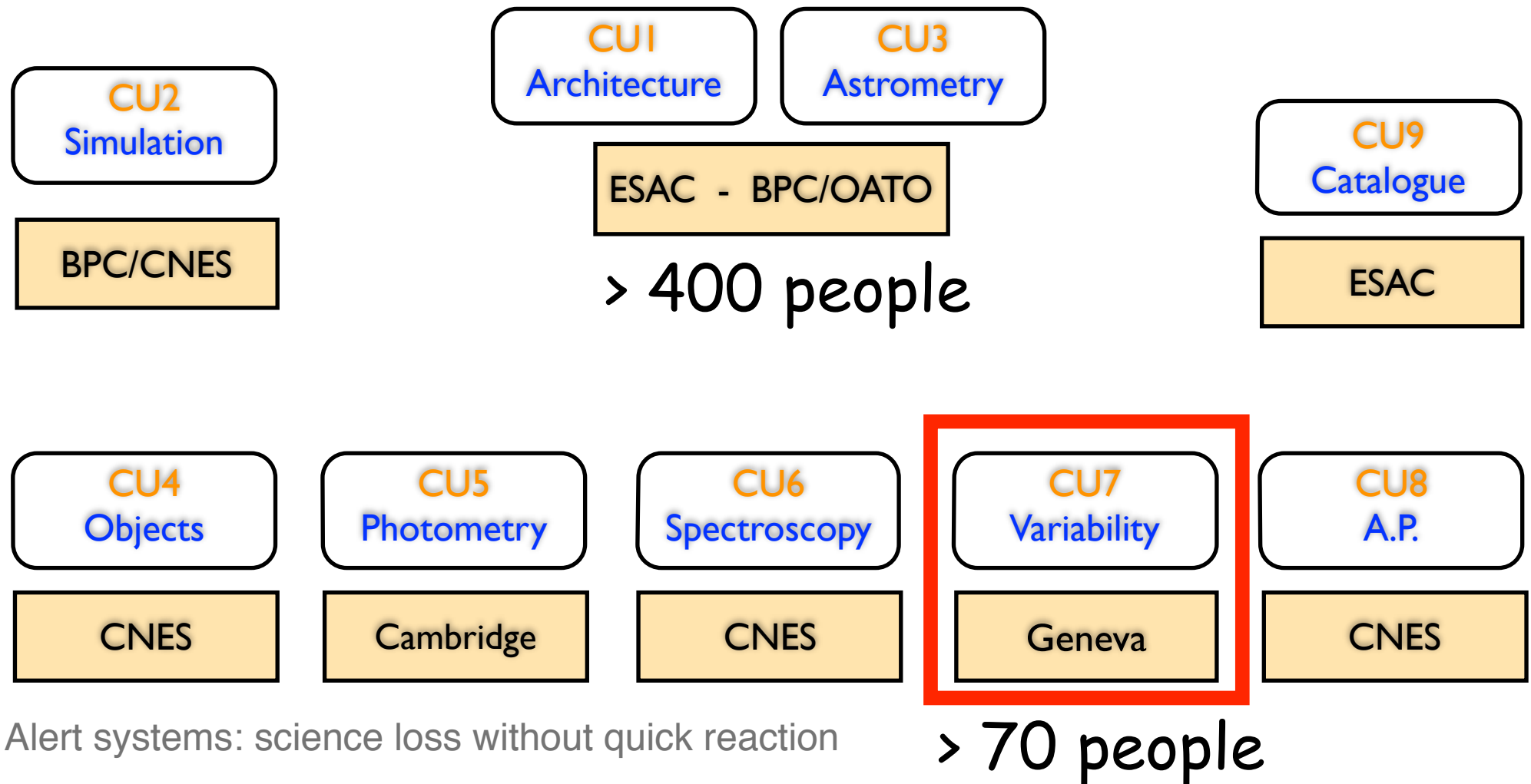
---

Two concepts: 1. Coordination Units (CU)  
2. Data Processing Centers (DPC)



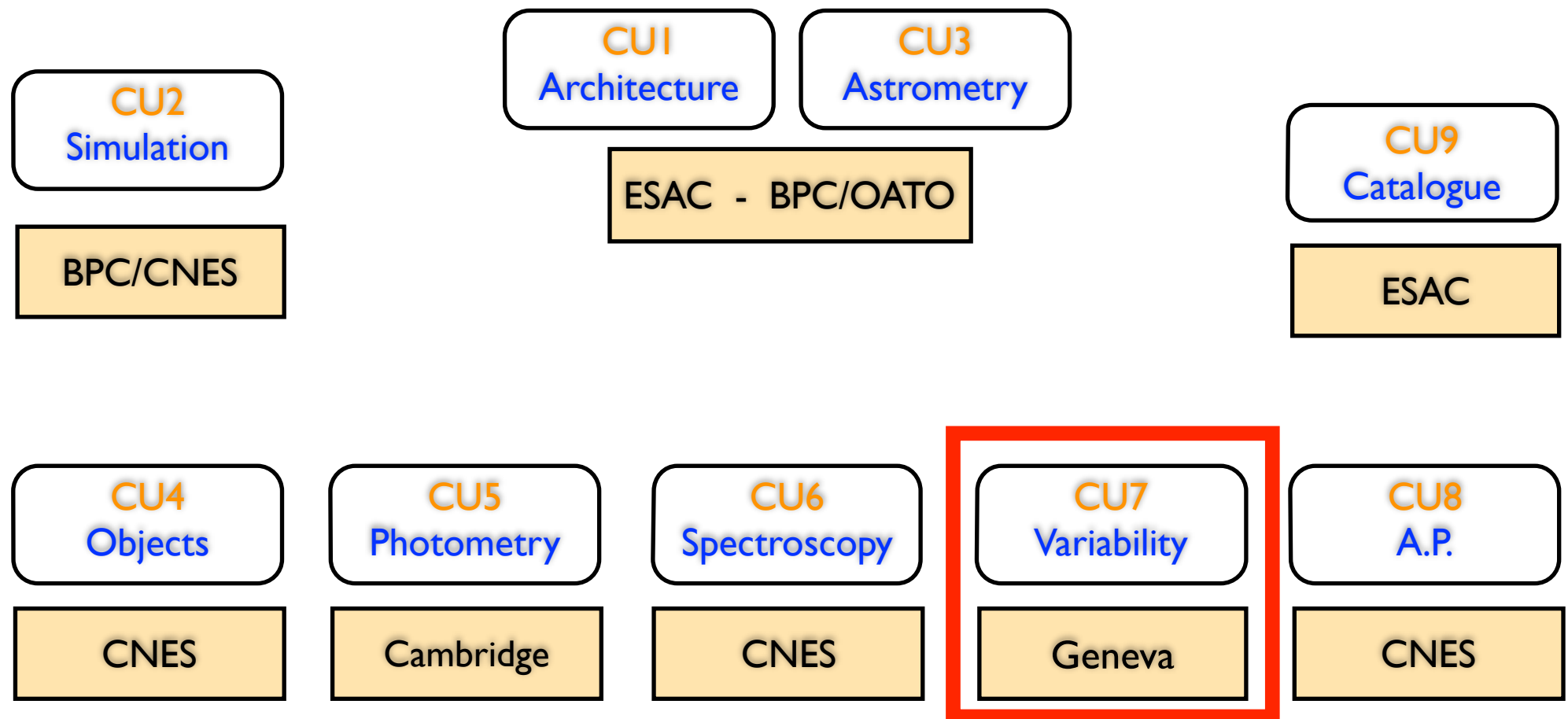
# Gaia Data Processing and Analysis Consortium DPAC

Two concepts: 1. Coordination Units (CU)  
2. Data Processing Centers (DPC)



# Gaia Data Processing and Analysis Consortium DPAC

Two concepts: 1. Coordination Units (CU)  
2. Data Processing Centers (DPC)

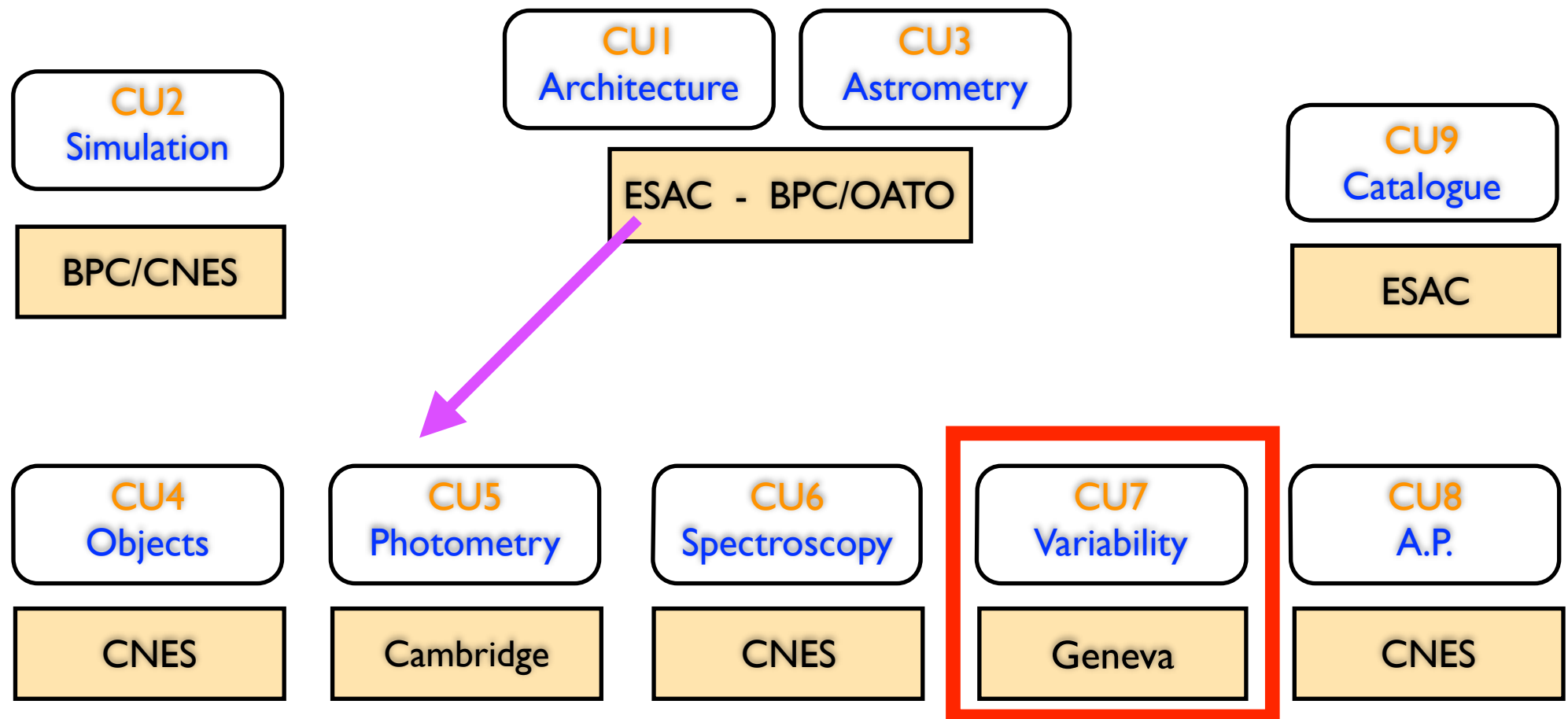


Alert systems: science loss without quick reaction

> 70 people

# Gaia Data Processing and Analysis Consortium DPAC

Two concepts: 1. Coordination Units (CU)  
2. Data Processing Centers (DPC)

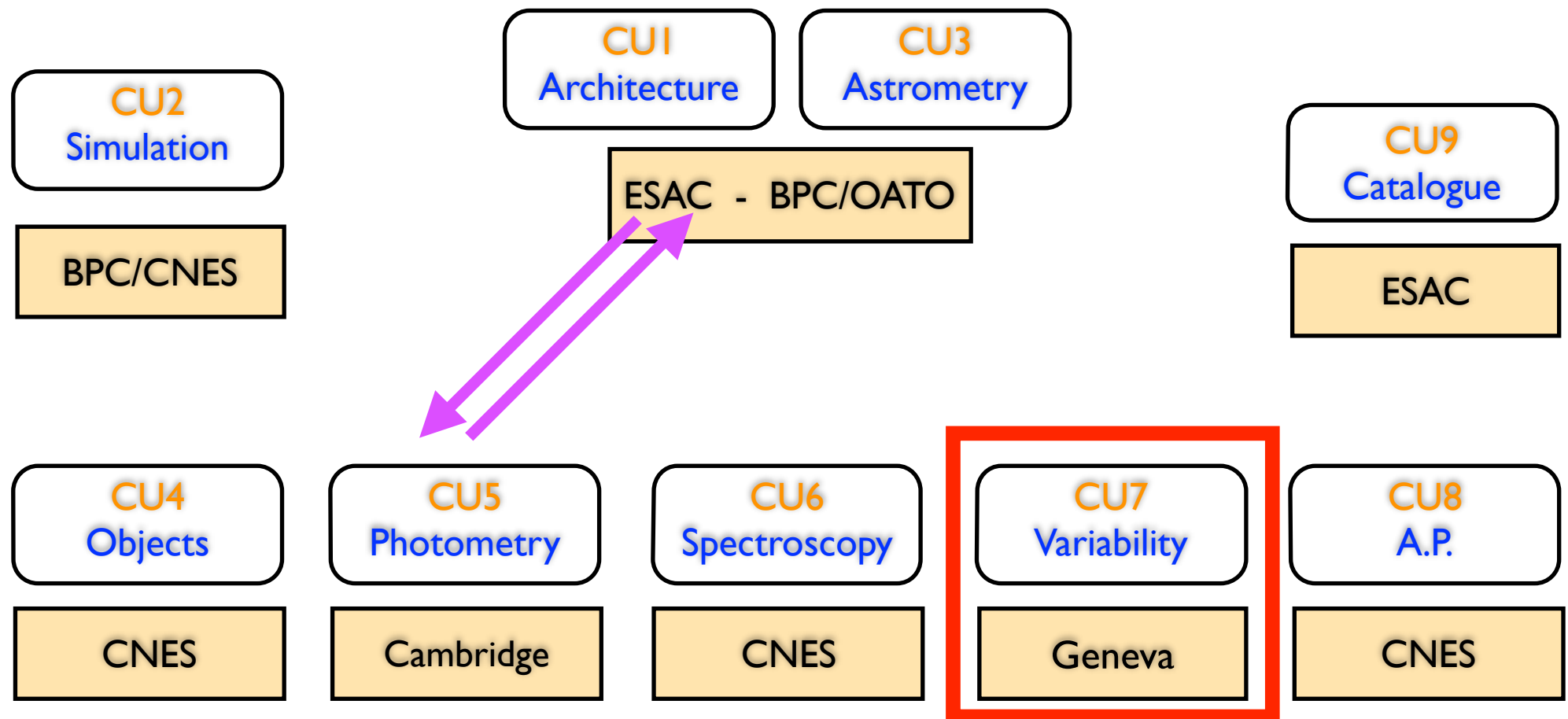


Alert systems: science loss without quick reaction

> 70 people

# Gaia Data Processing and Analysis Consortium DPAC

Two concepts: 1. Coordination Units (CU)  
2. Data Processing Centers (DPC)



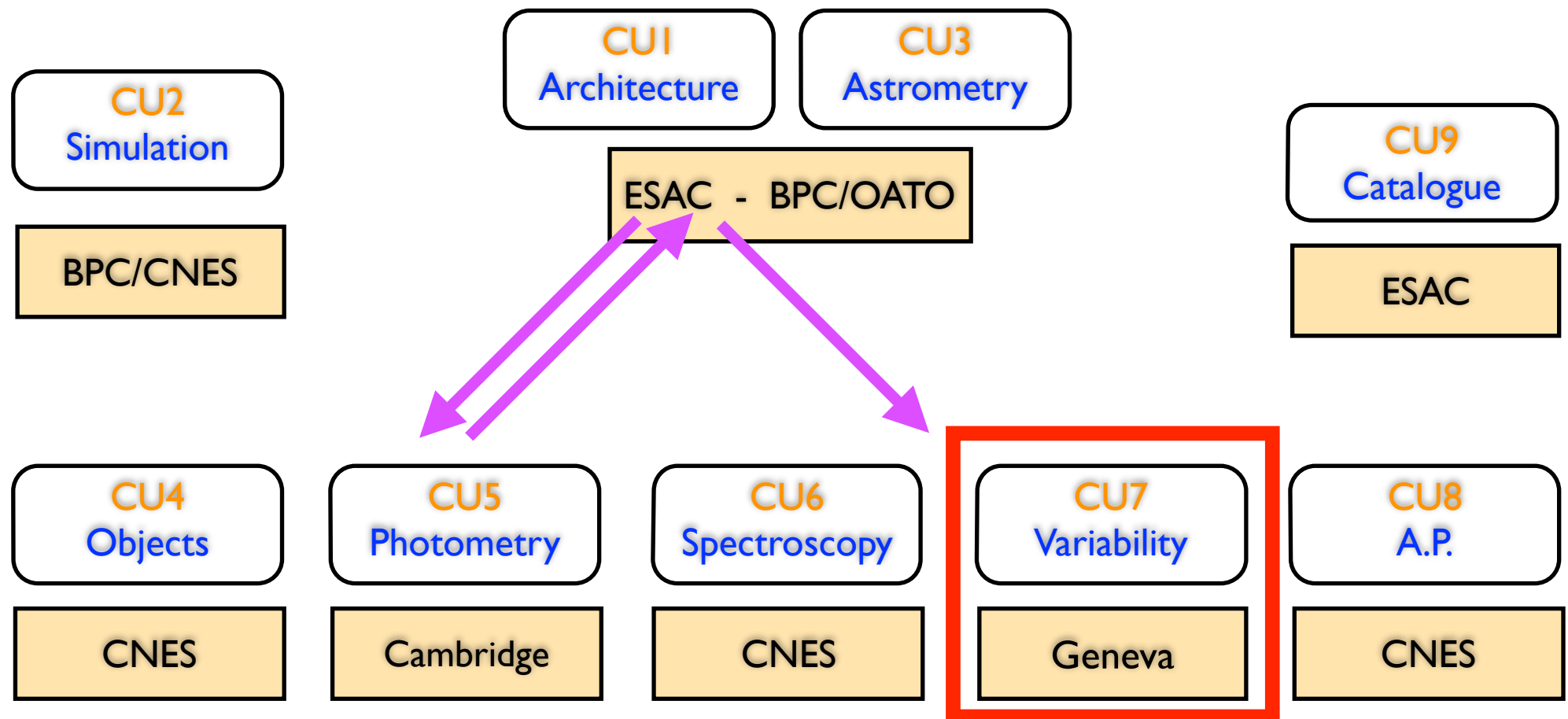
Alert systems: science loss without quick reaction

> 70 people



# Gaia Data Processing and Analysis Consortium DPAC

Two concepts: 1. Coordination Units (CU)  
2. Data Processing Centers (DPC)



Alert systems: science loss without quick reaction

> 70 people

# Operation Rehearsals

---

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

# Operation Rehearsals

---

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

Variability Processing and Analysis is **not** “in operation”

# Operation Rehearsals

---

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

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

# 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)

# 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 now on

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

# 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 now on

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

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

# 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 now on

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

**This exercise/training is 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



# 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 now on

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

**This exercise/training is 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

# Operation Rehearsal (OR5 stage 2) on Gaia EPSL data

---

# Operation Rehearsal (OR5 stage 2) on Gaia EPSL data

---

Hooray — — First real Gaia data !!!!



# Operation Rehearsal (OR5 stage 2) on Gaia EPSL data

---

Hooray — — First real Gaia data !!!!



Very significant effort from a large fraction of CU7 members!

# The Operation Rehearsal data

---

# The Operation Rehearsal data

---

**Data set** { 28 days of Ecliptic Scanning law  
3 days of Nominal Scanning law

# 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)

# 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





# 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)



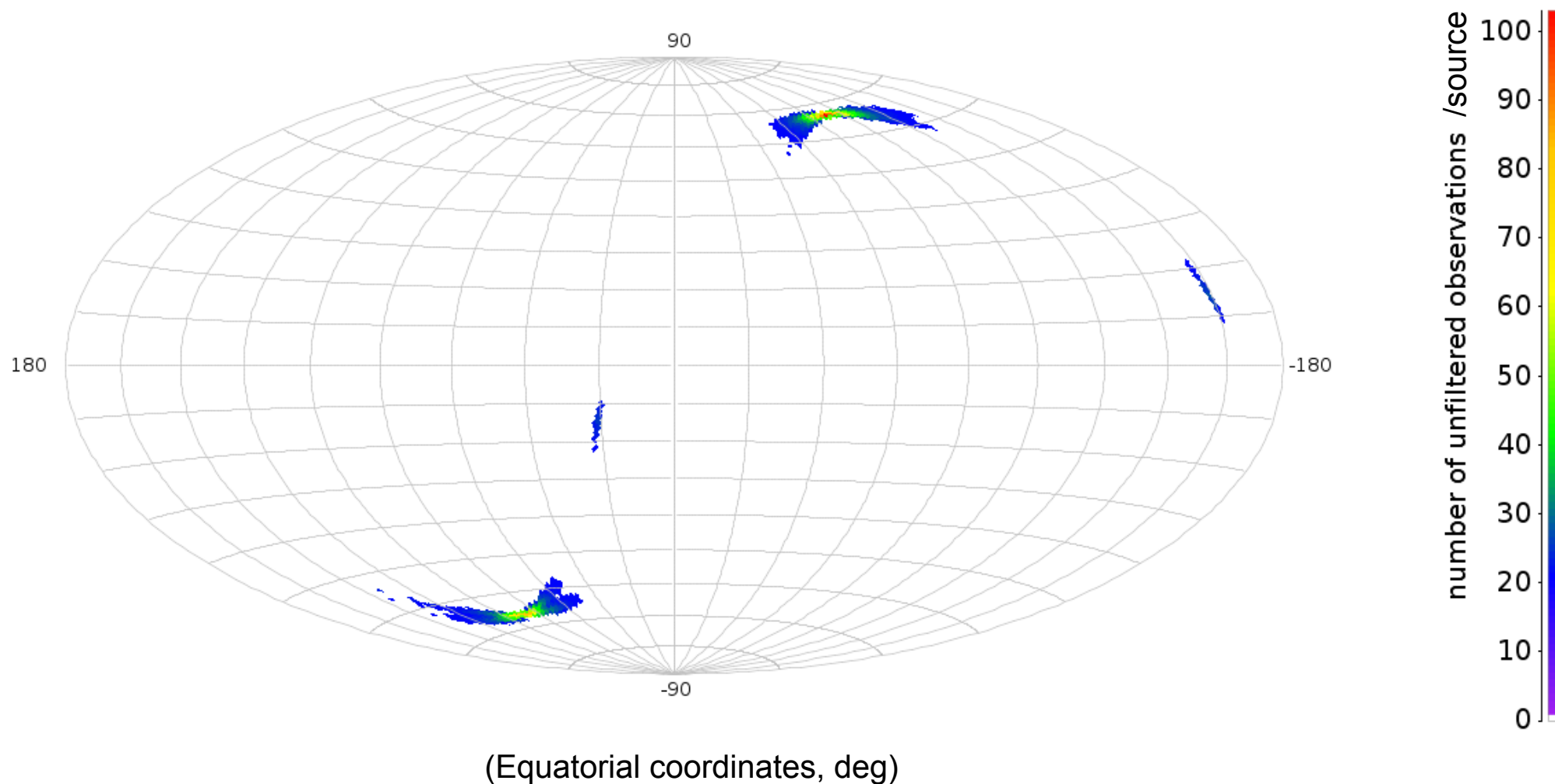
**Selection:**  
20 measurements in either G, BP, RP  
No repeated observations within 100 minutes



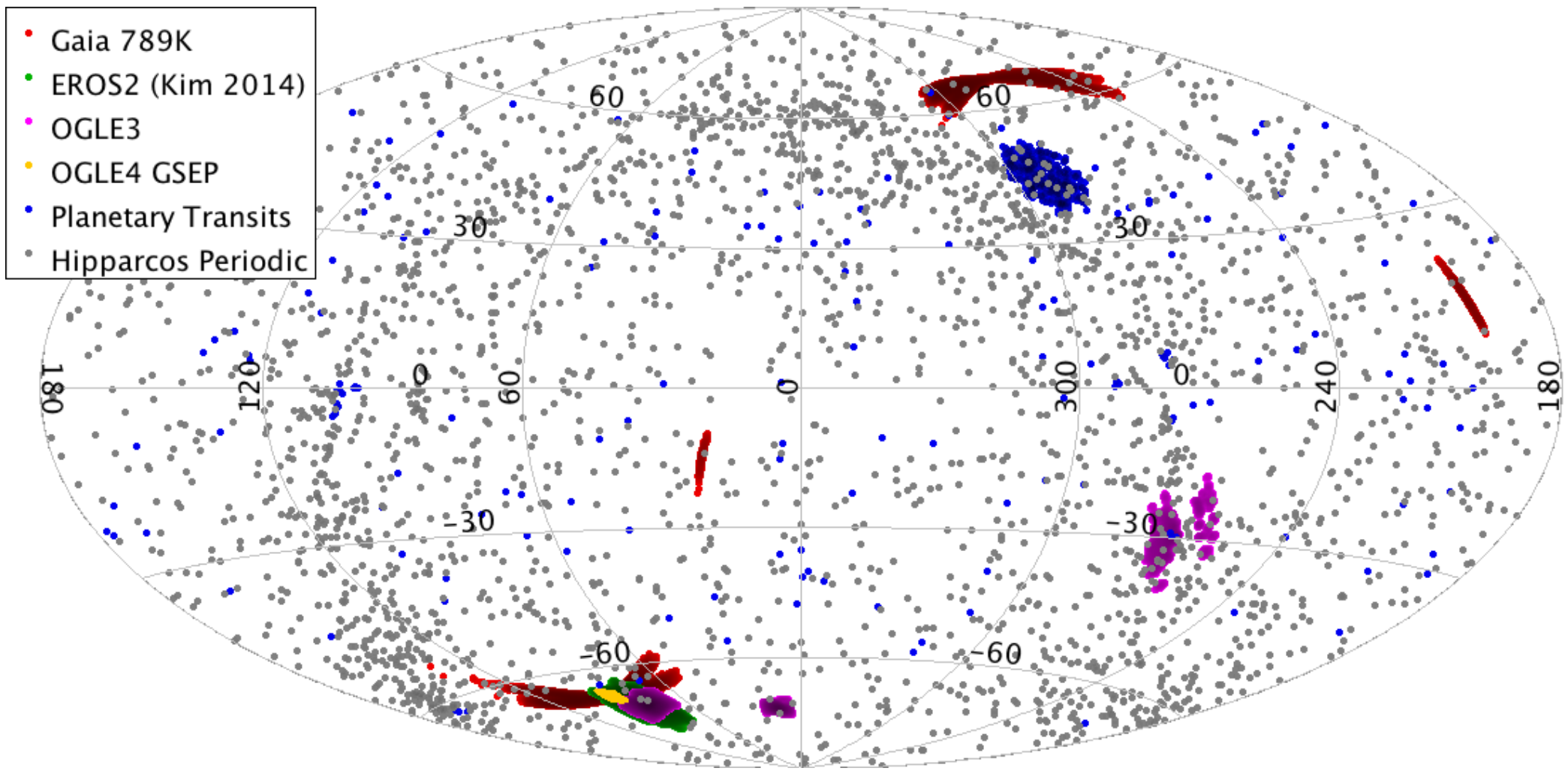
**790,000 sources**

# Mean number of obs/source (789K)

Map of number of unfiltered observations (equatorial coordinates, 1pix = 0.84 deg<sup>2</sup>)  
Catalog GAIA-OR5S2-NO-REPEAT-GT20FOV, band

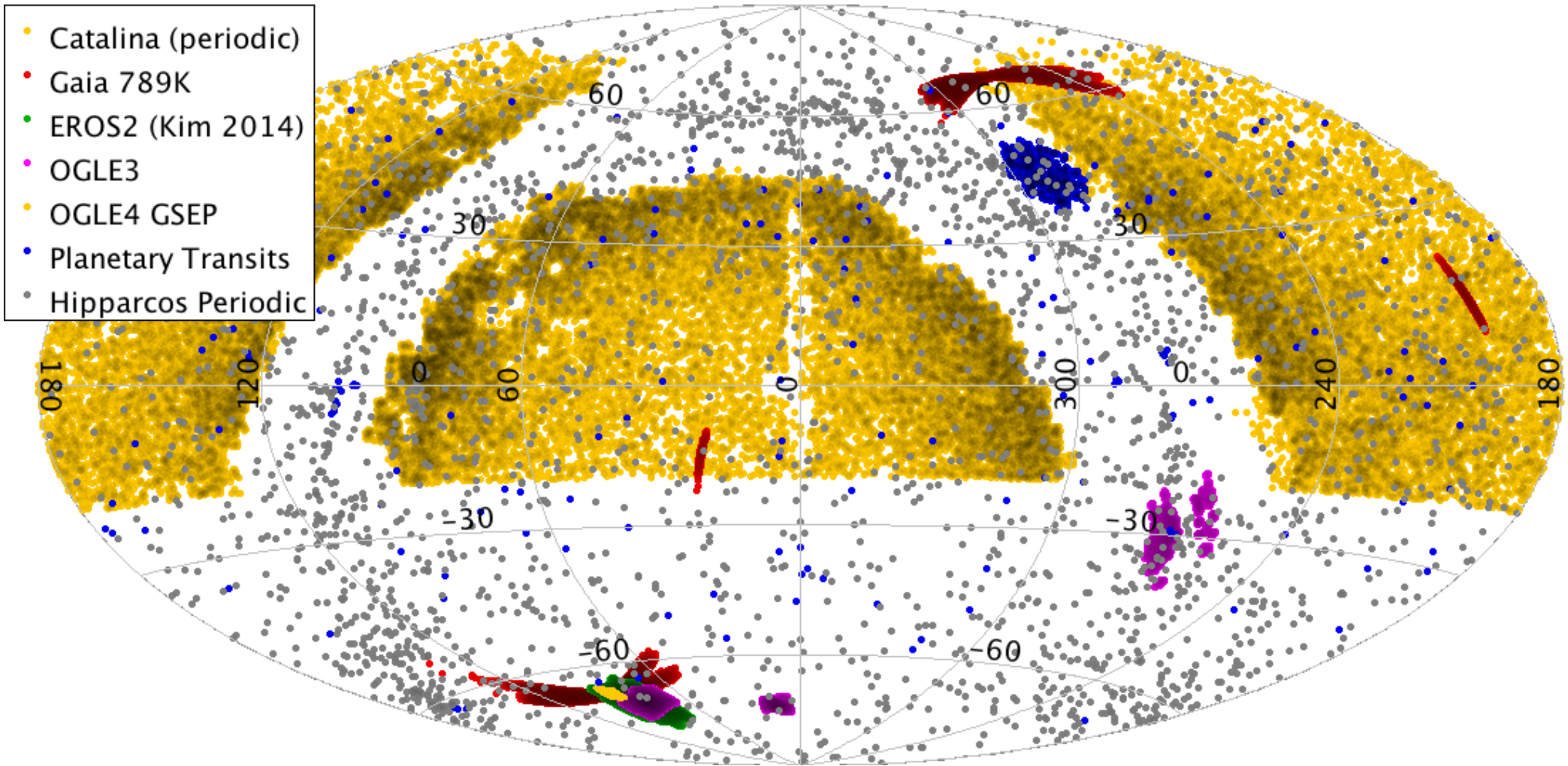


# Gaia 789K + EROS2, OGLE, Hipparcos, Planetary Transits



(Equatorial coordinates, deg)

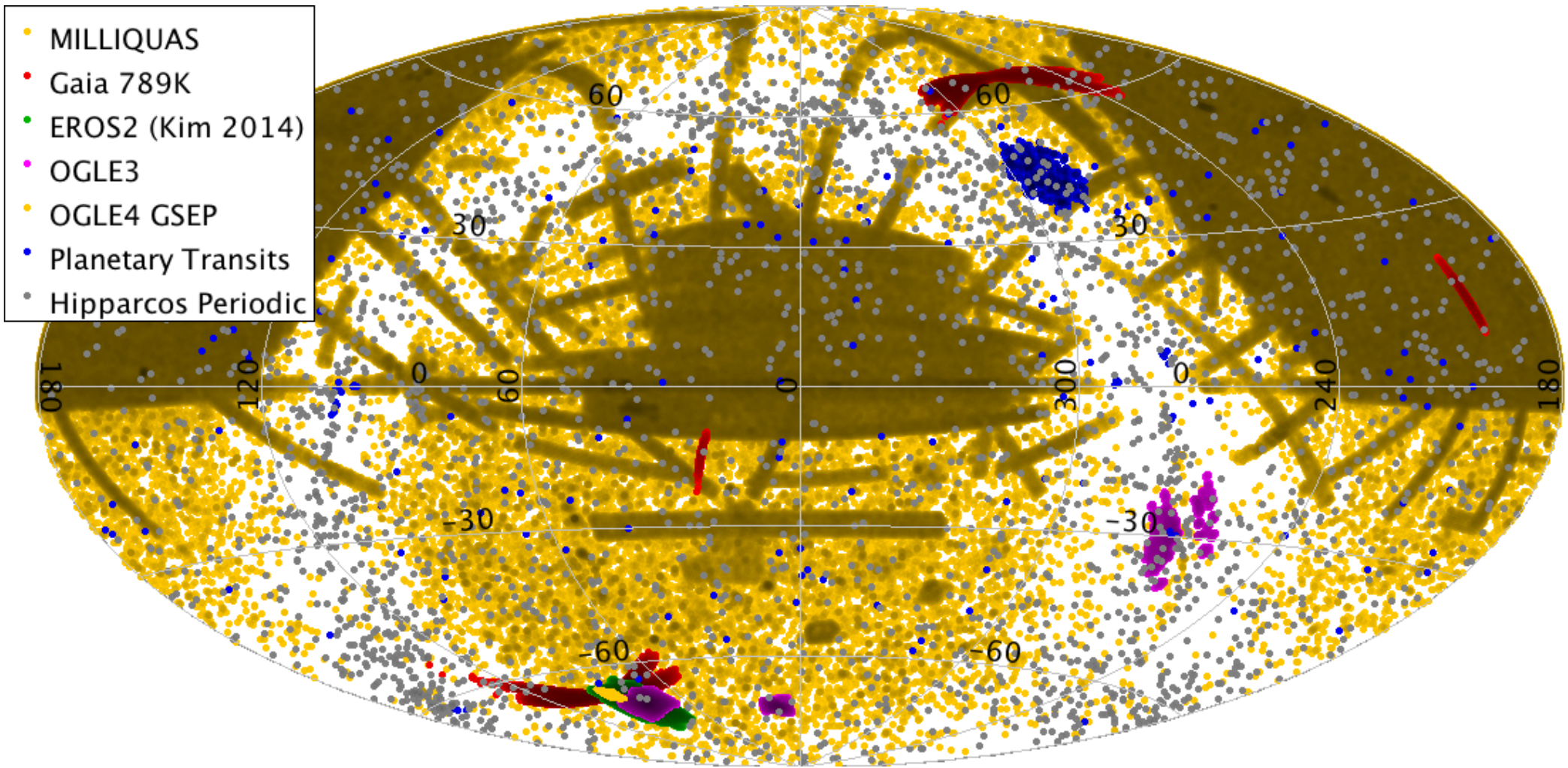
# + Catalina (periodic)



(Equatorial coordinates, deg)

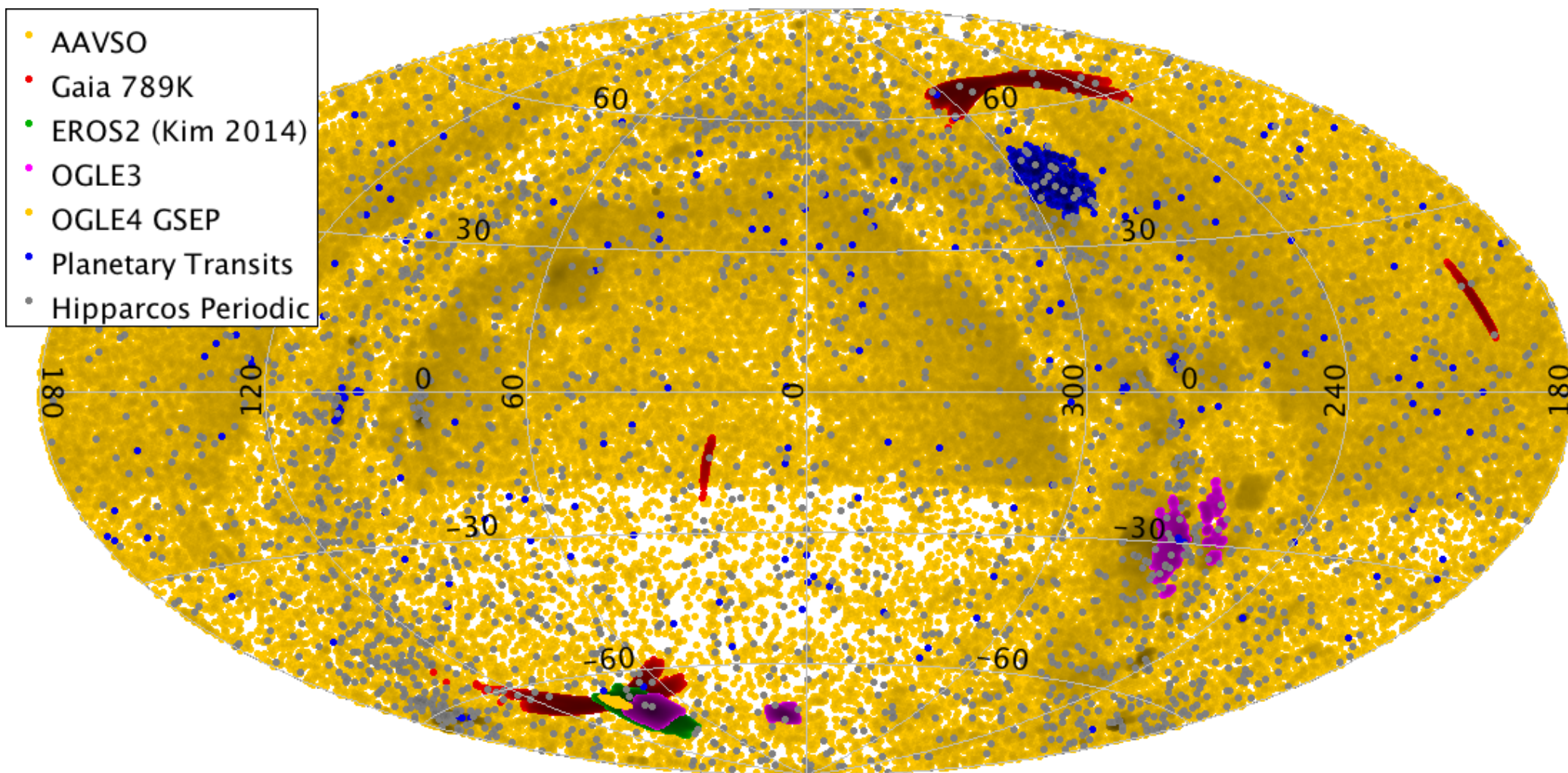


# + MILLIQUAS



(Equatorial coordinates, deg)

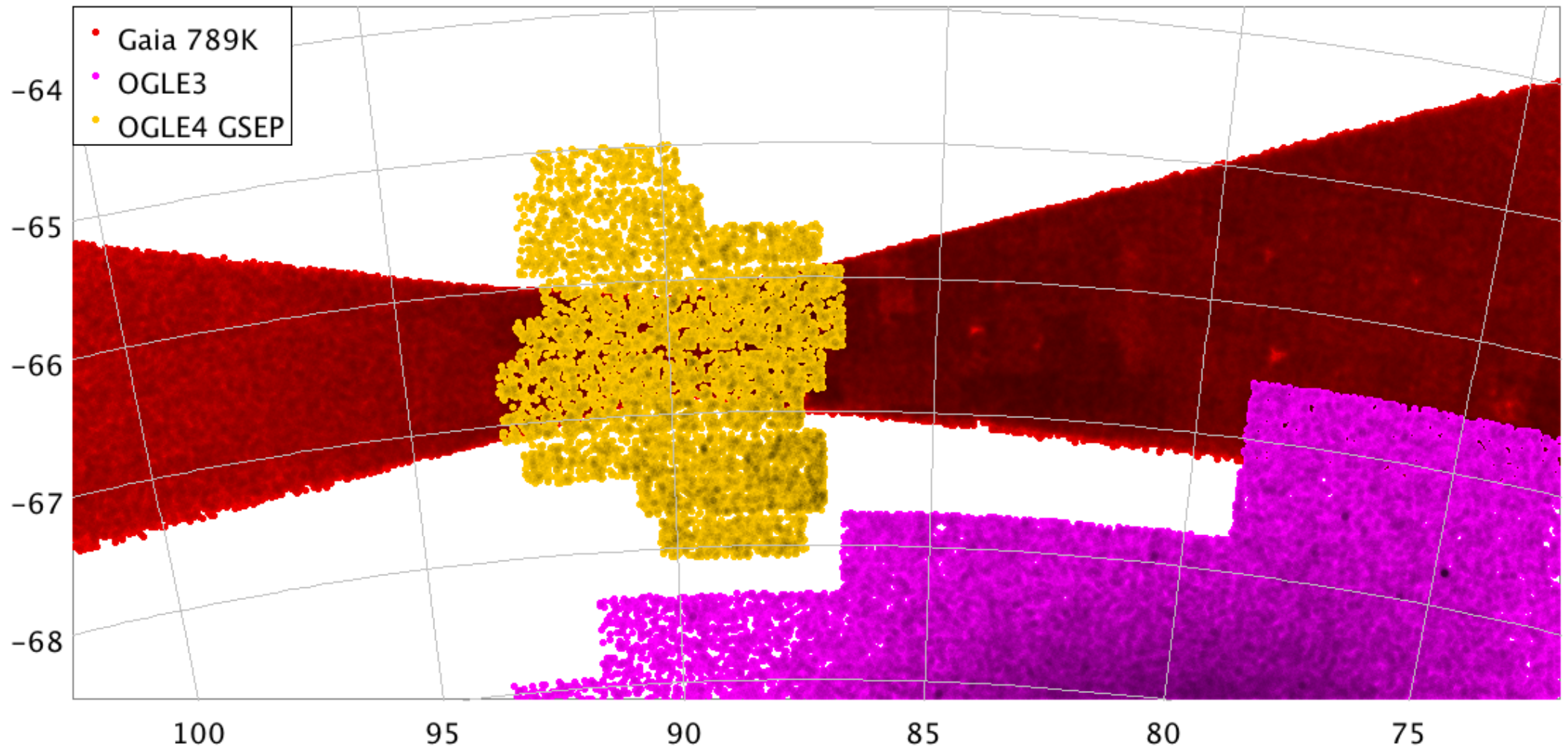
# + AAVSO



(Equatorial coordinates, deg)

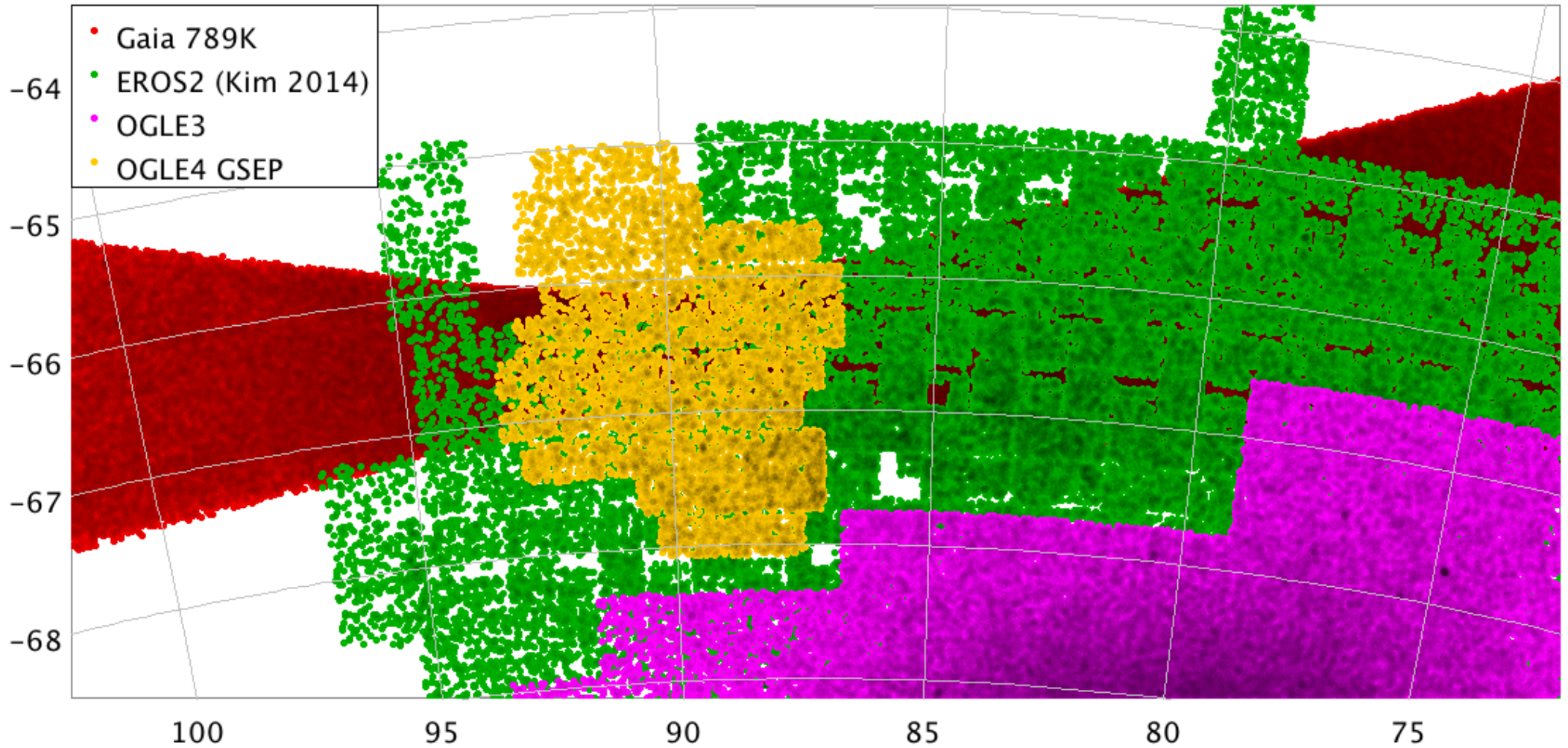


# SEP: Gaia + OGLE



(Equatorial coordinates, deg)

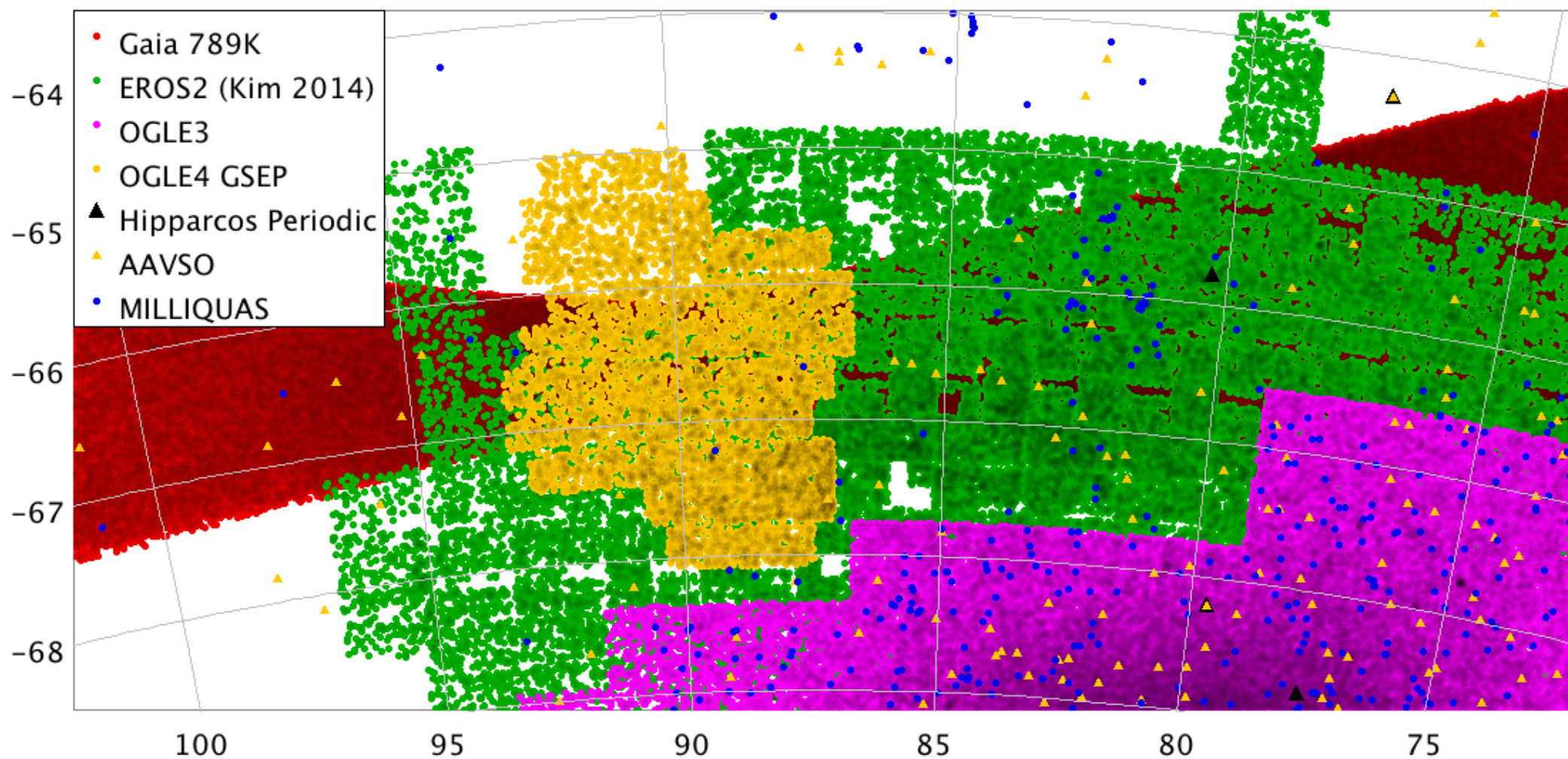
# SEP: Gaia + OGLE, EROS2



(Equatorial coordinates, deg)

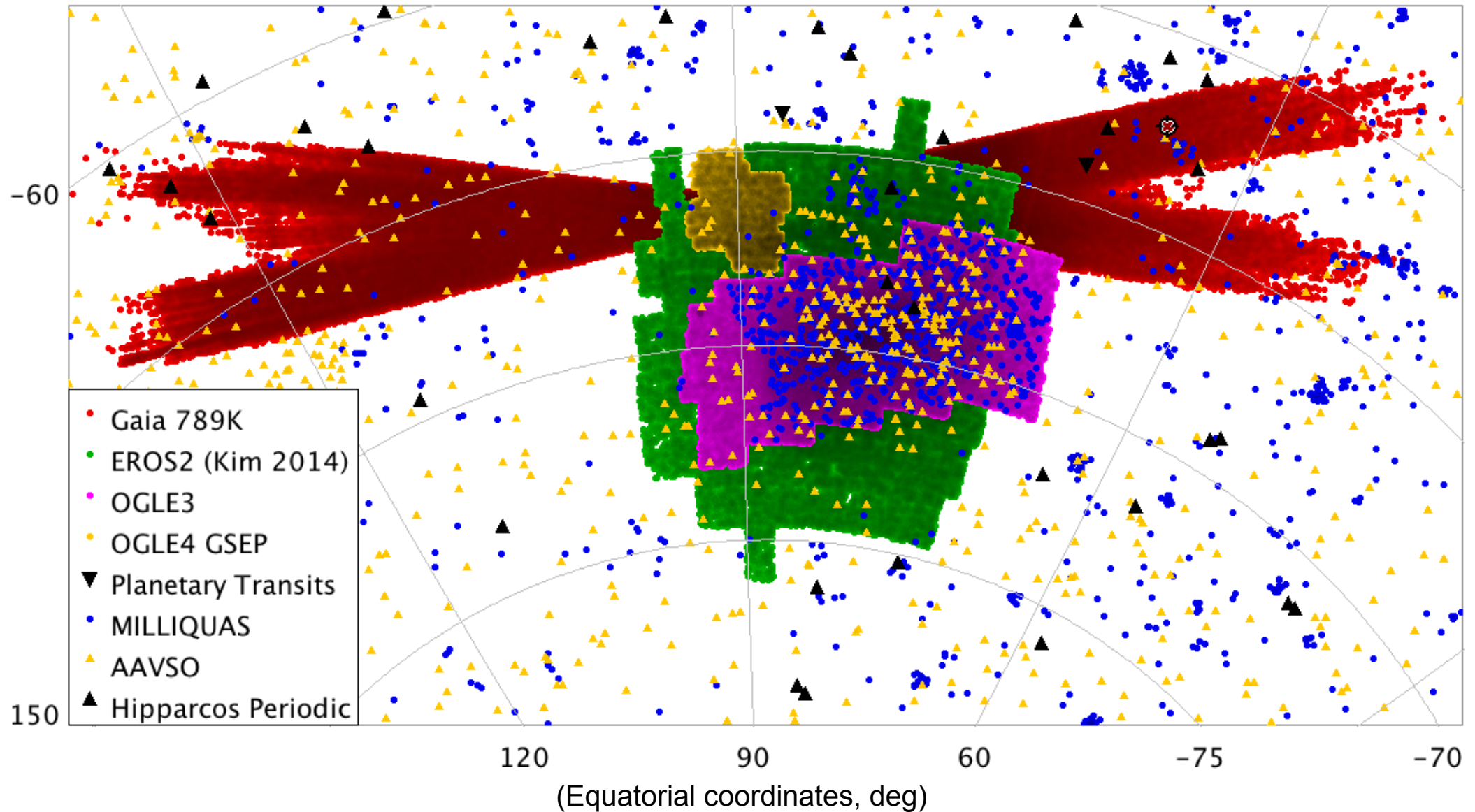


# SEP: Gaia + OGLE, EROS2, Hipparcos, AAVSO, MILLIQUAS



(Equatorial coordinates, deg)

# SEP: Gaia + OGLE, EROS2, Hipparcos, AAVSO, MILLIQUAS, Planetary Transits



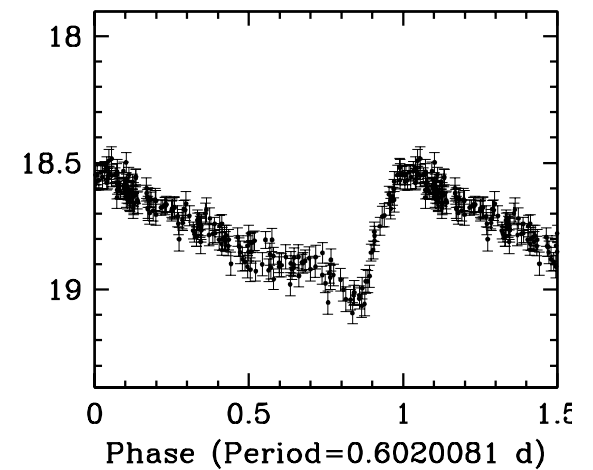
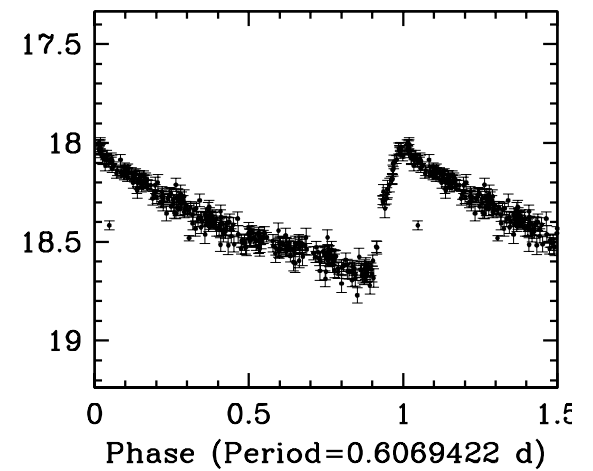
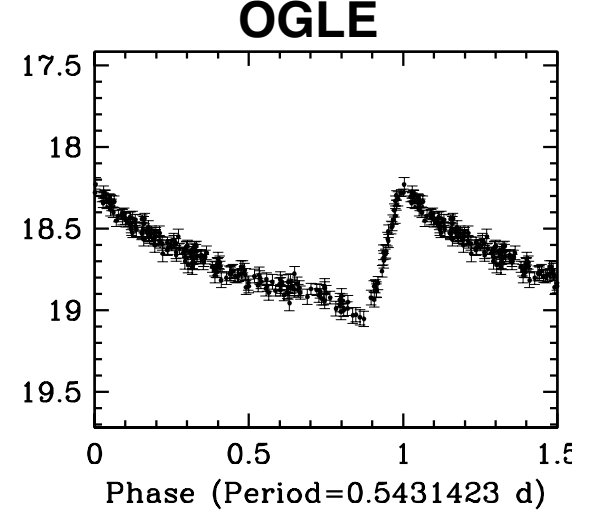
# To get the data flavour

## Comparison with OGLE

---

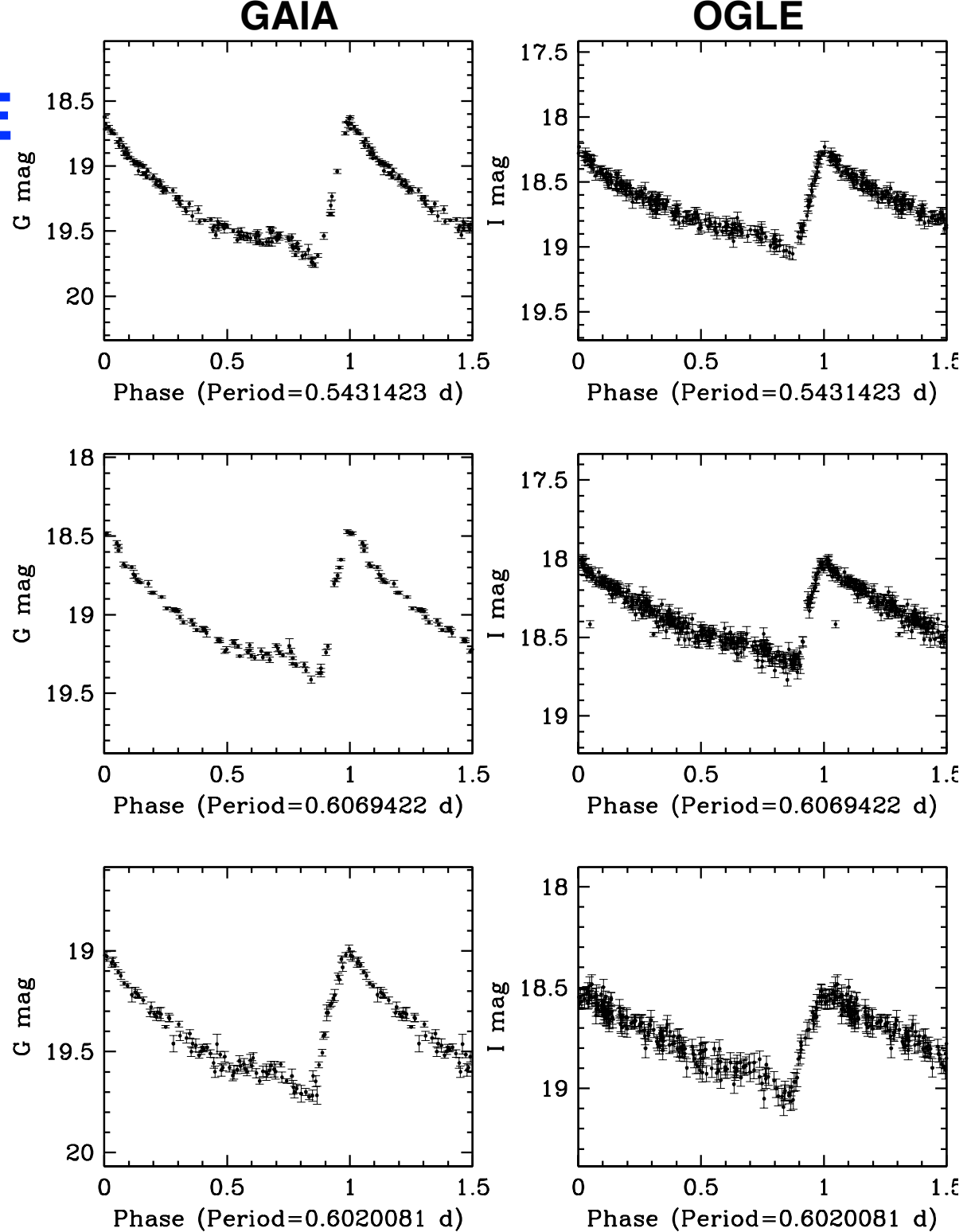
Image of the Week (March 05):  
RR Lyrae stars

*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.*



# To get the data flavour Comparison with OGLE

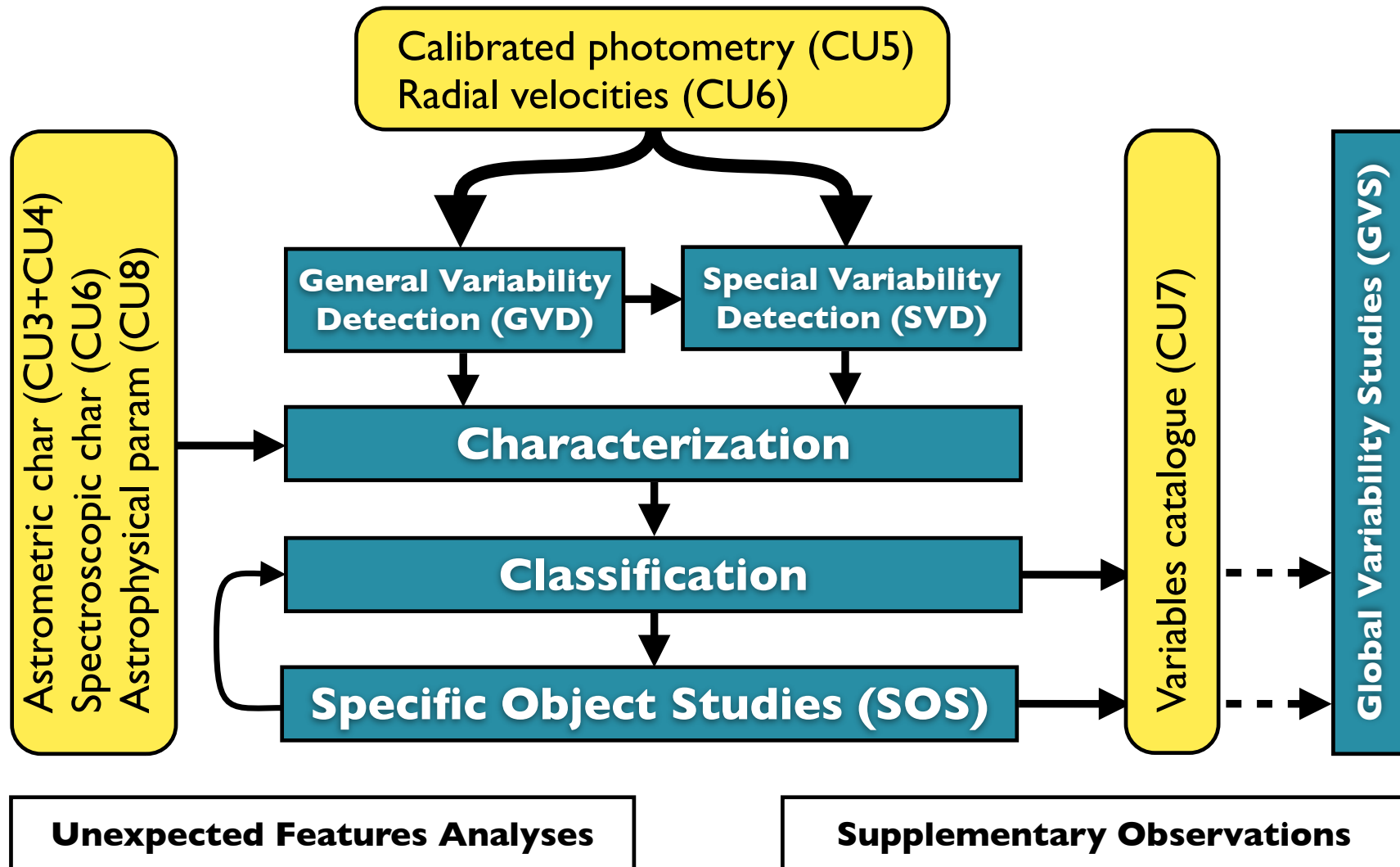
Image of the Week (March 05):  
RR Lyrae stars



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

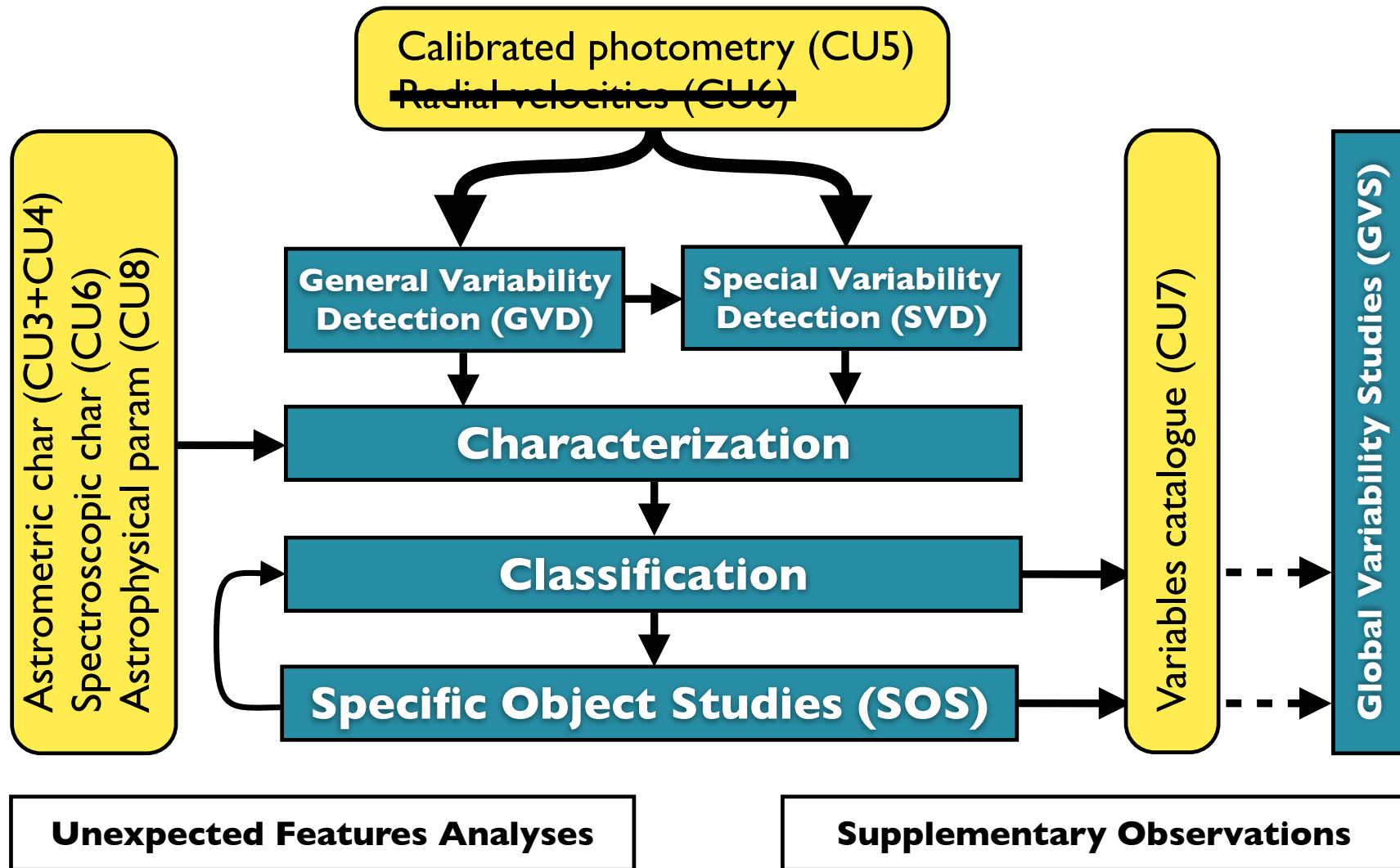
# CU7 / DPCG Variability Analysis

As presented at EWASS meeting in Geneva in 2014



# CU7 / DPCG Variability Analysis

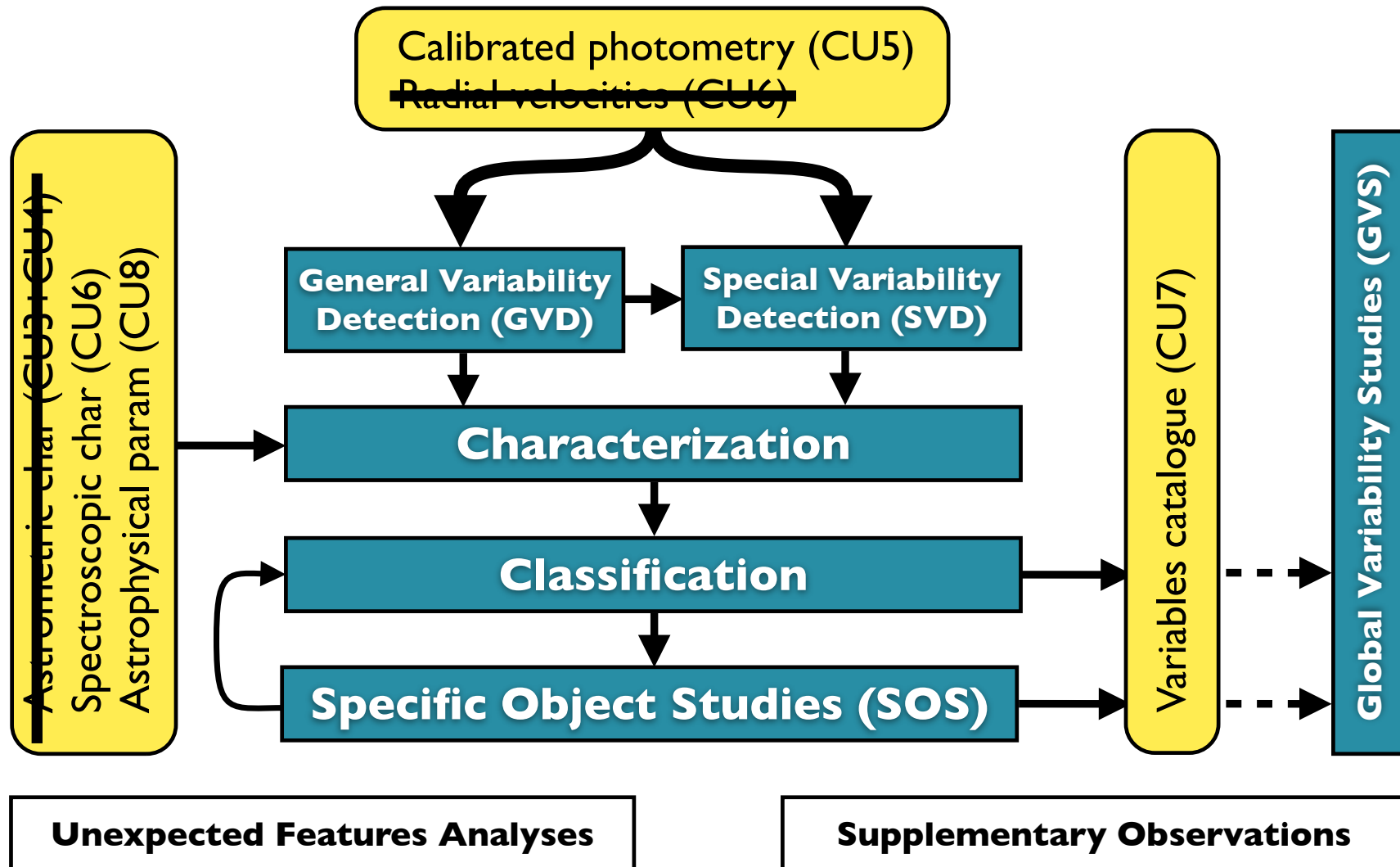
As presented at EWASS meeting in Geneva in 2014





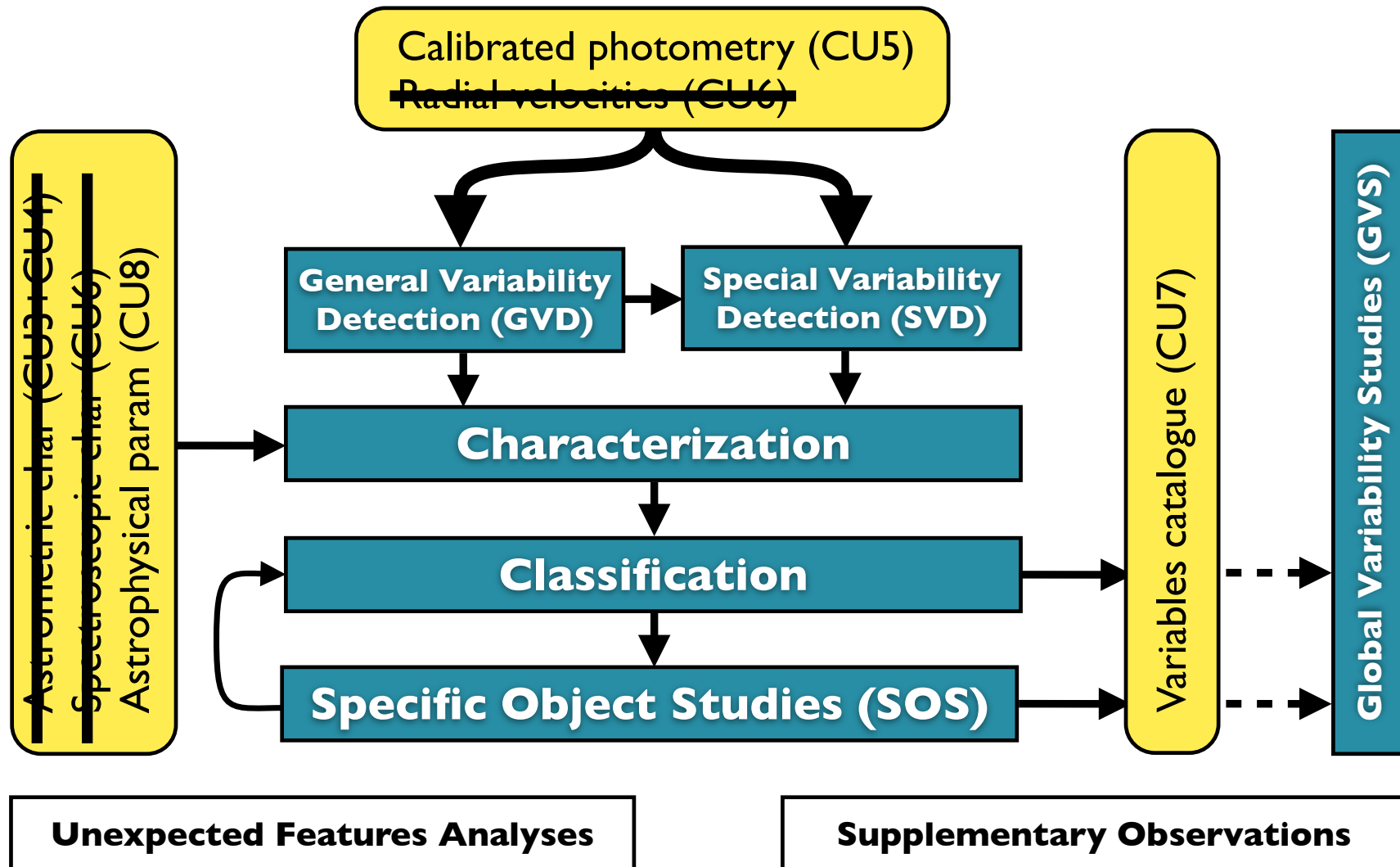
# CU7 / DPCG Variability Analysis

As presented at EWASS meeting in Geneva in 2014



# CU7 / DPCG Variability Analysis

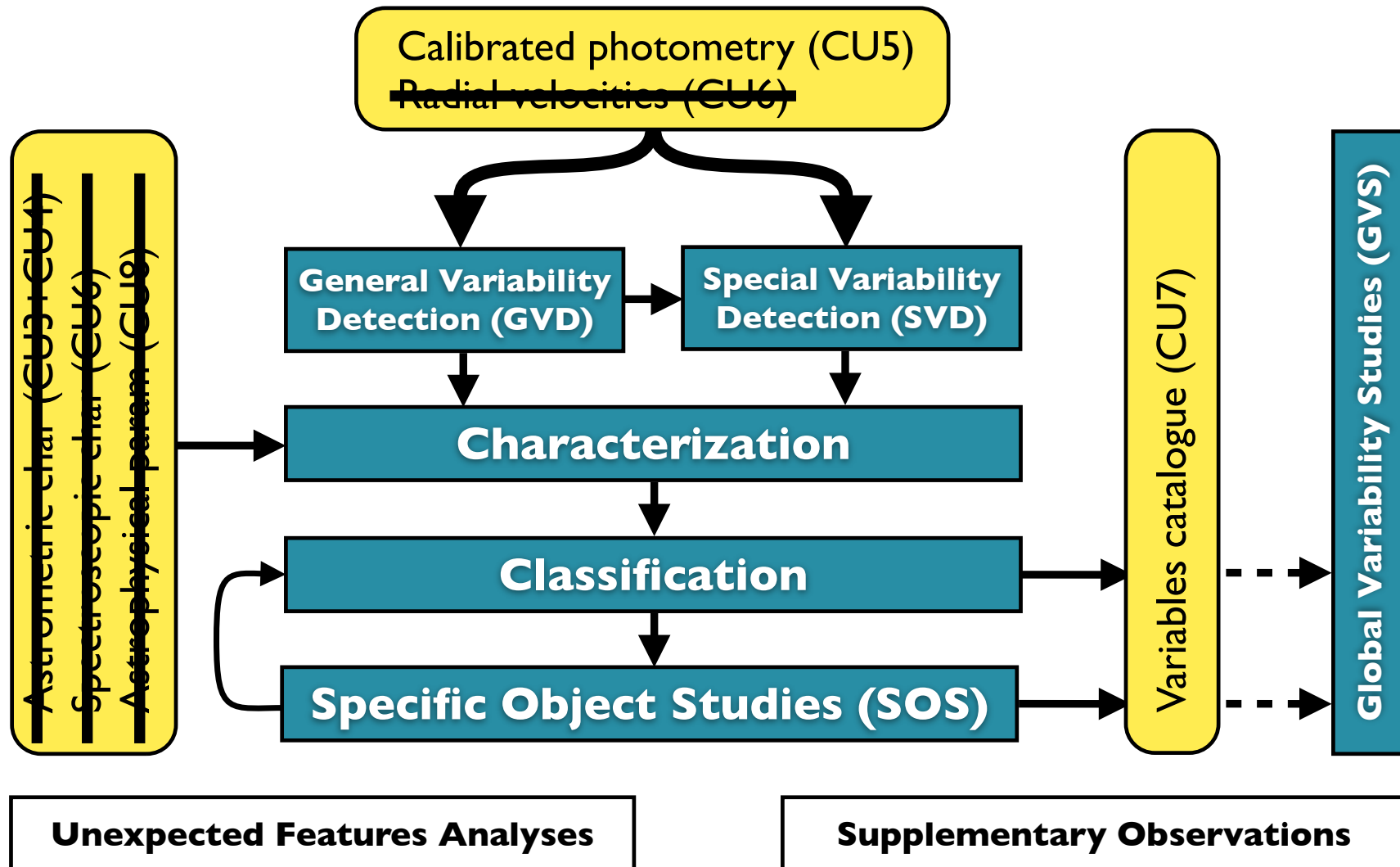
As presented at EWASS meeting in Geneva in 2014





# CU7 / DPCG Variability Analysis

As presented at EWASS meeting in Geneva in 2014



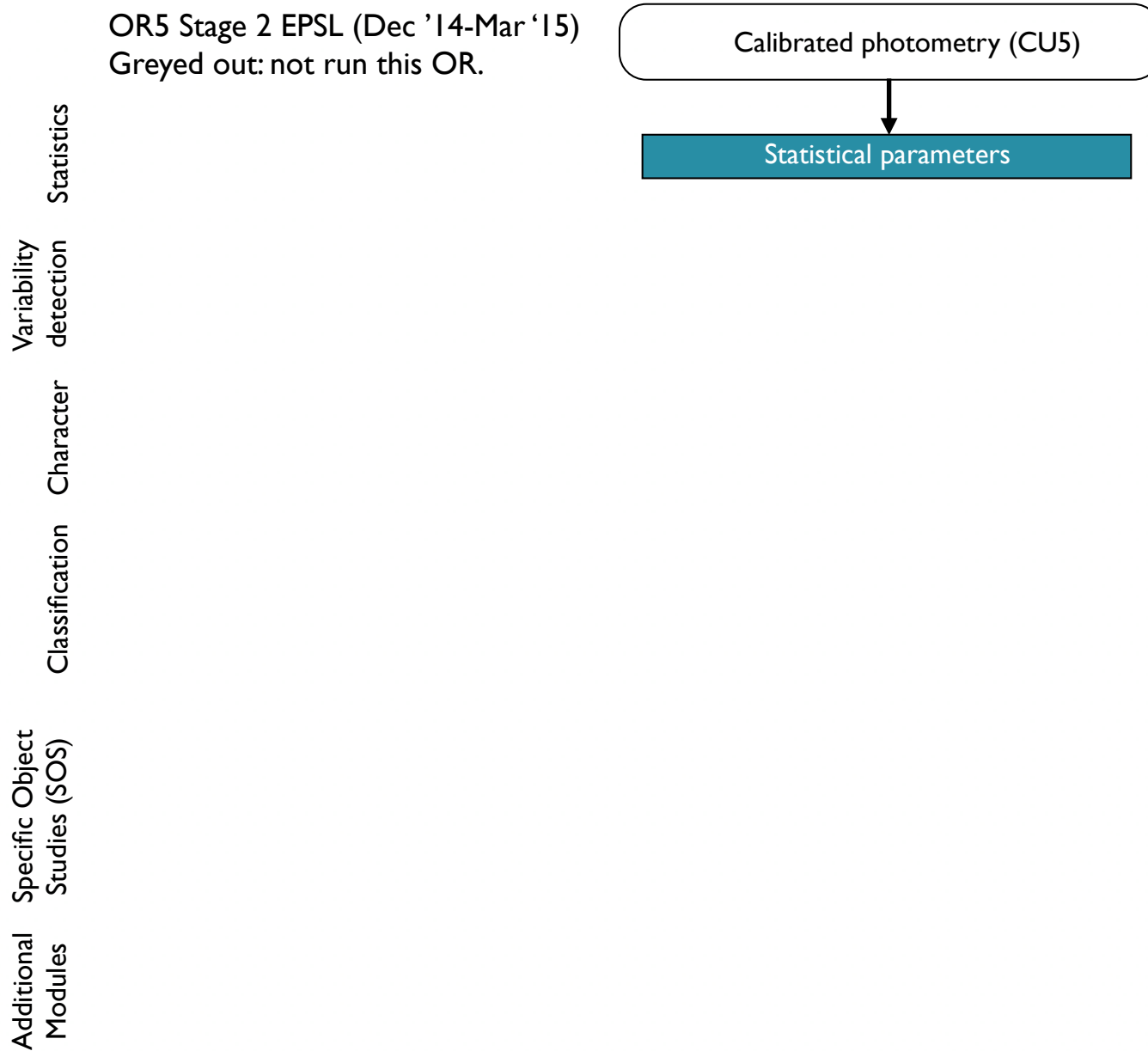
# CU7 processing chain

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

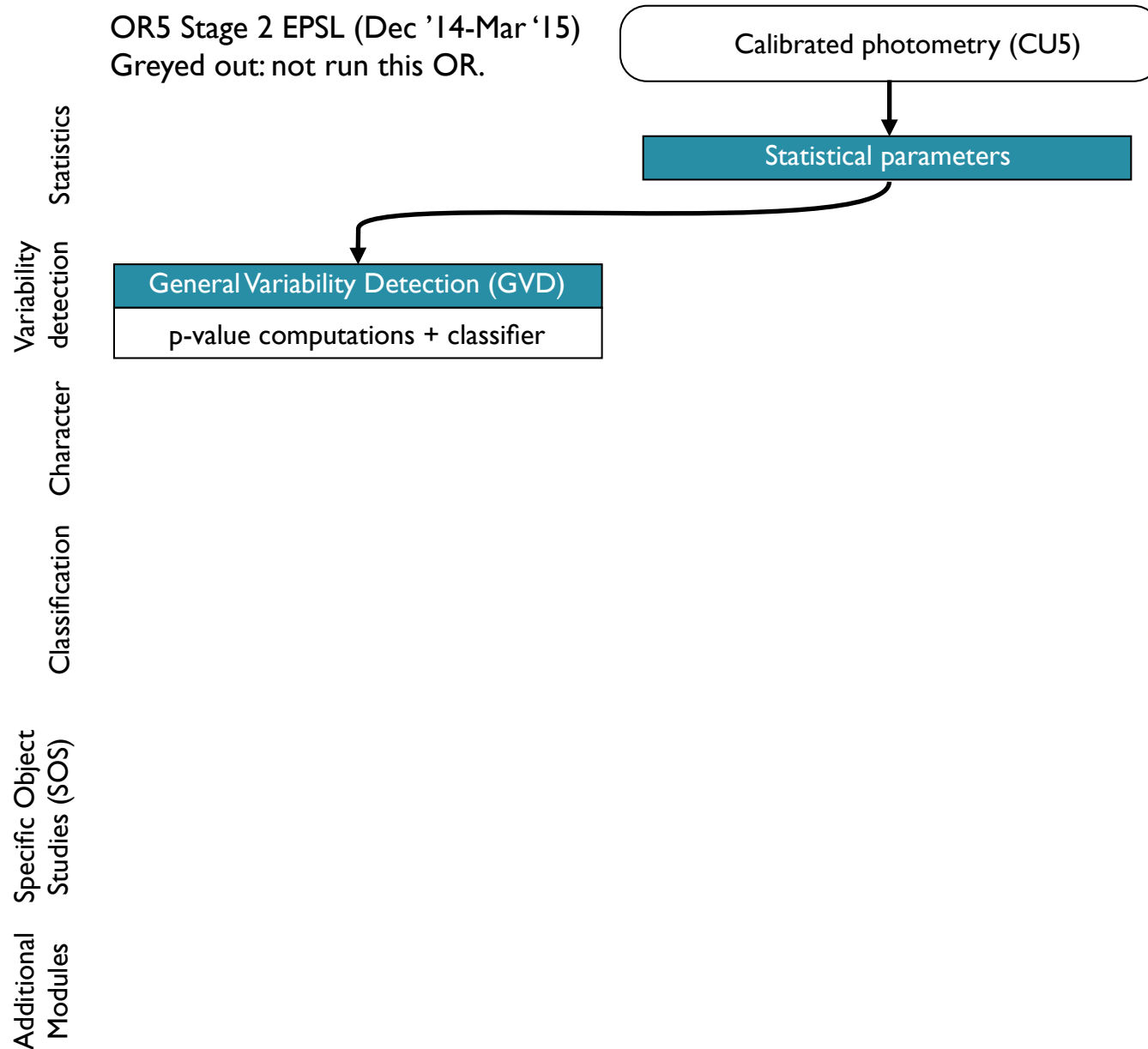
Calibrated photometry (CU5)

Additional Modules  
Specific Object Studies (SOS)  
Classification  
Character  
Variability detection  
Statistics

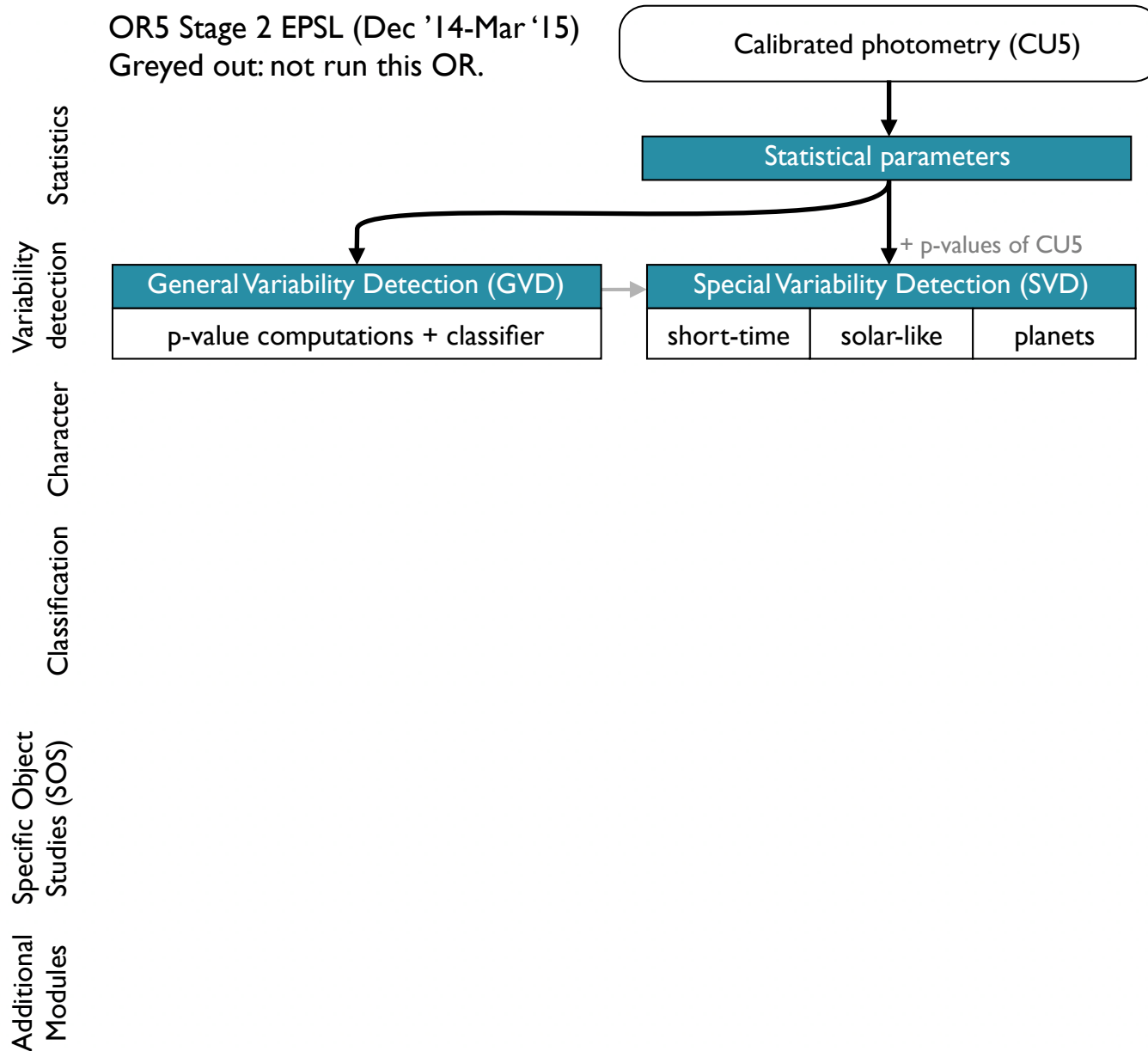
# CU7 processing chain



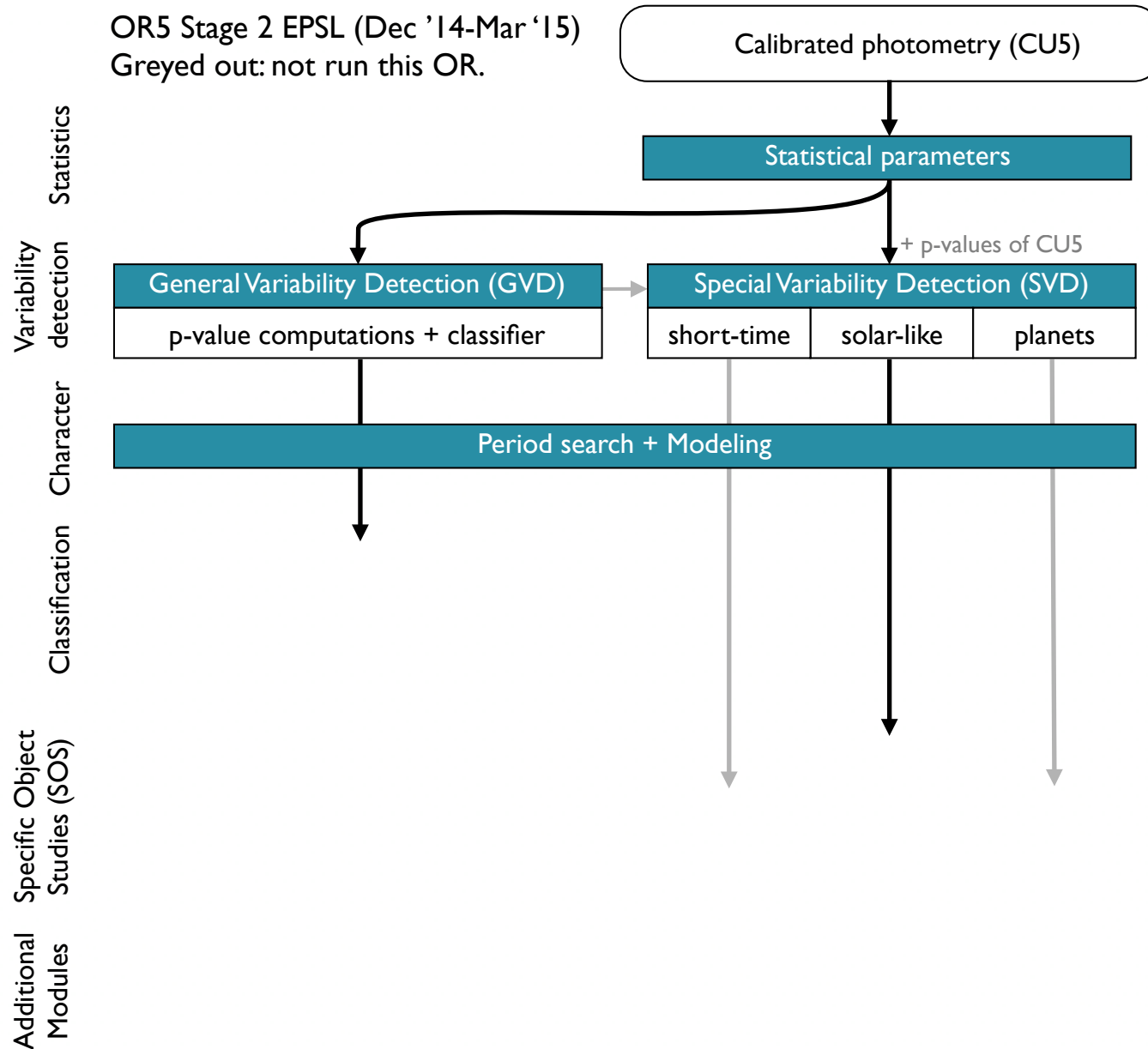
# CU7 processing chain



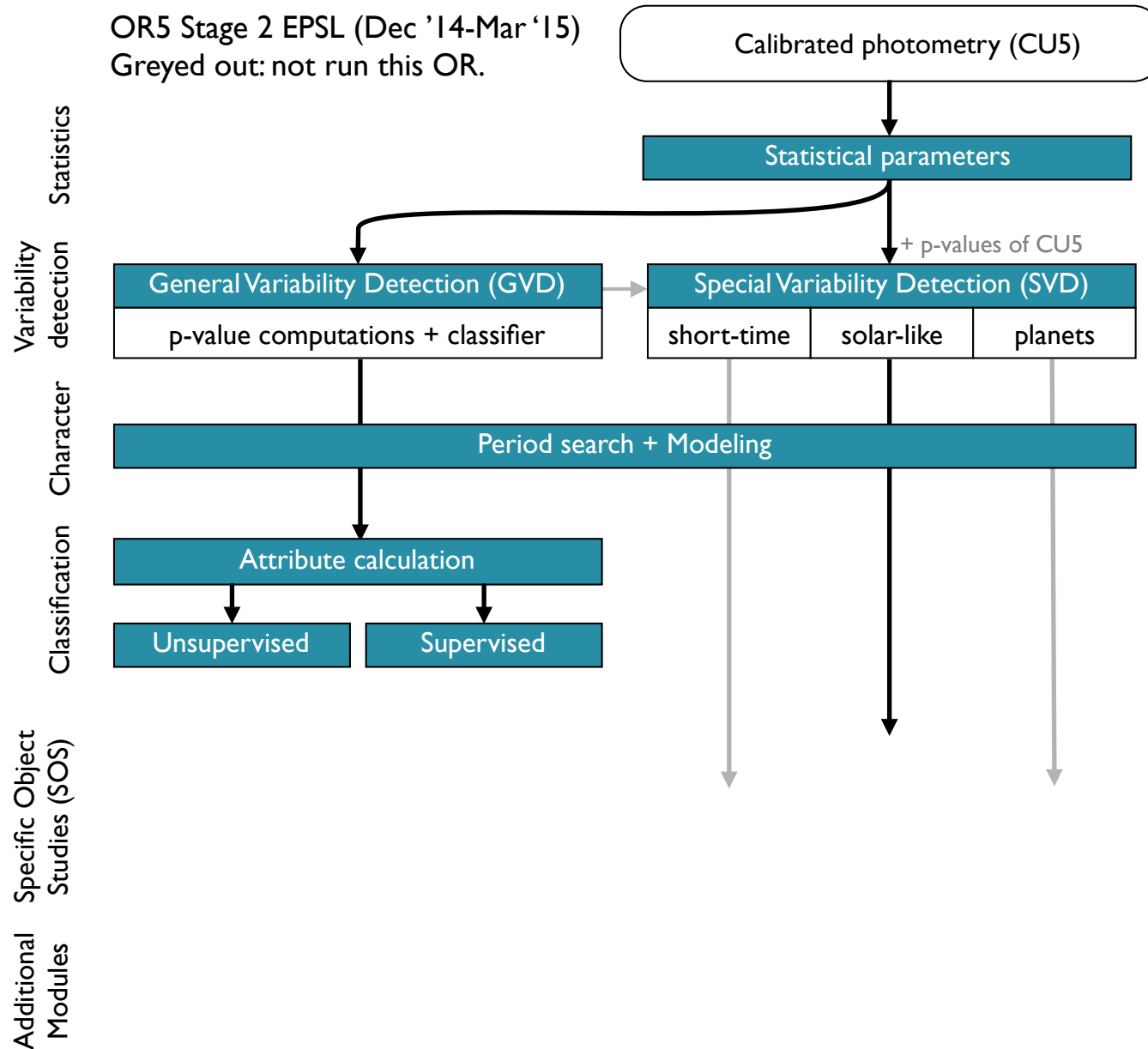
# CU7 processing chain



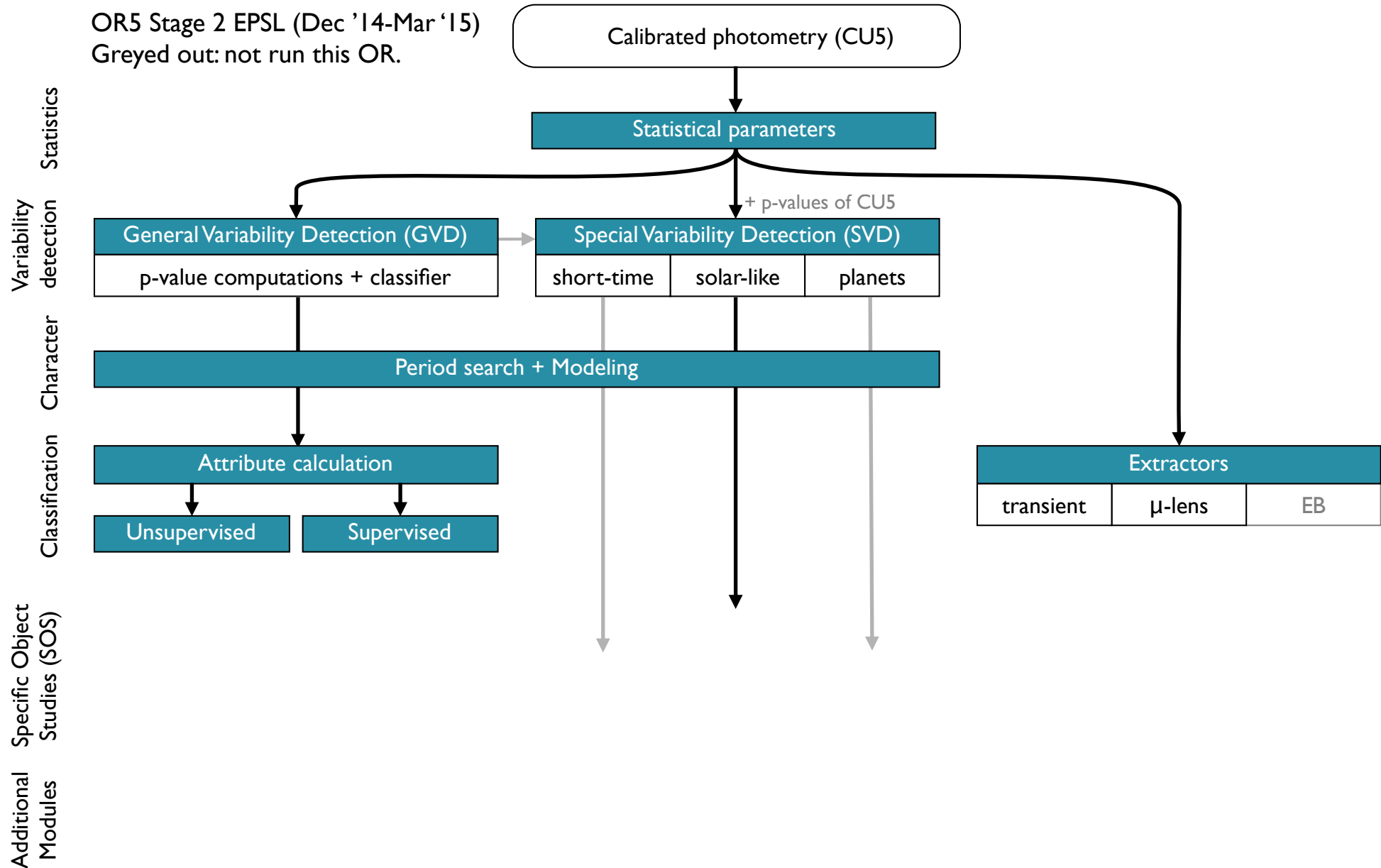
# CU7 processing chain



# CU7 processing chain

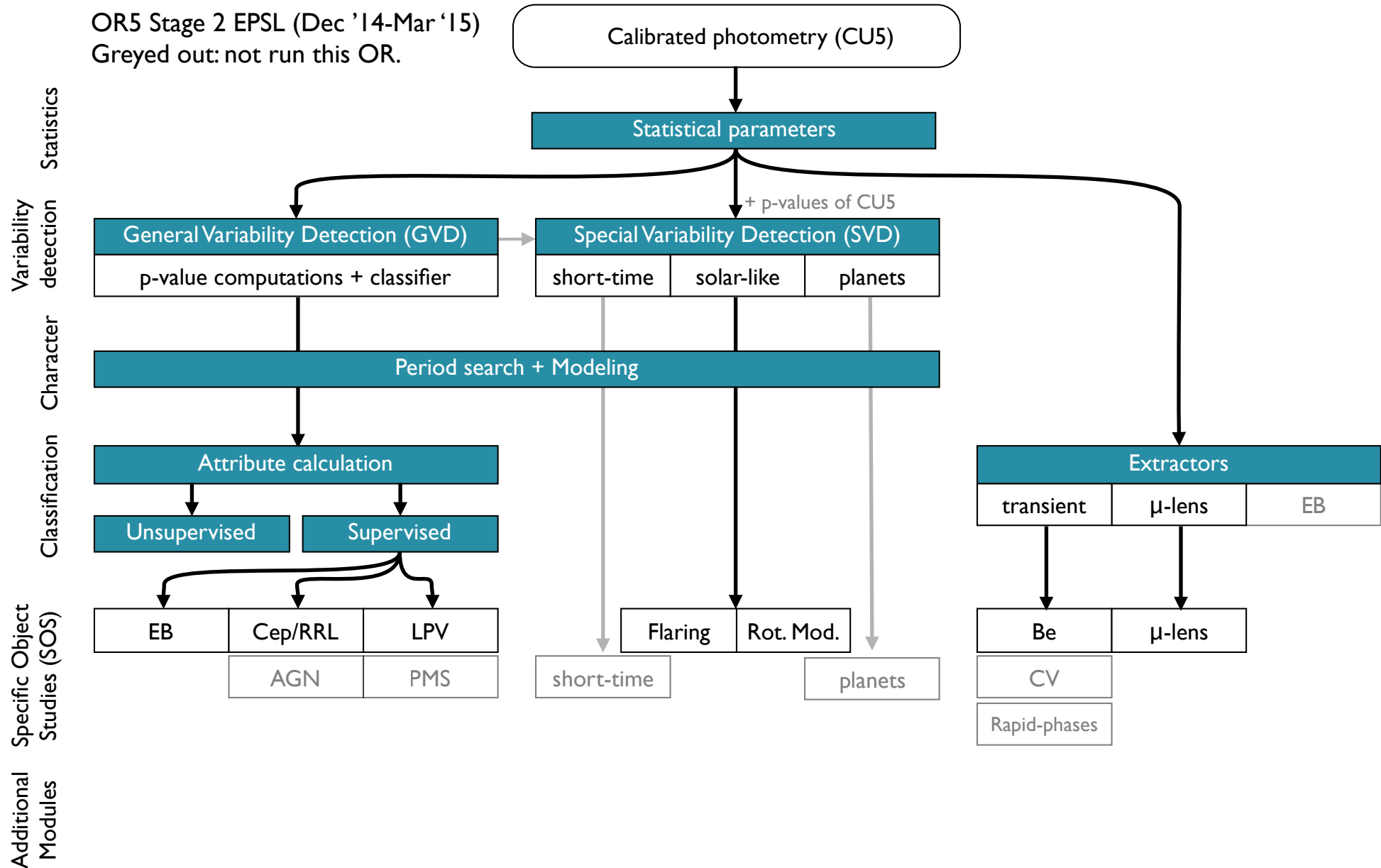


# CU7 processing chain

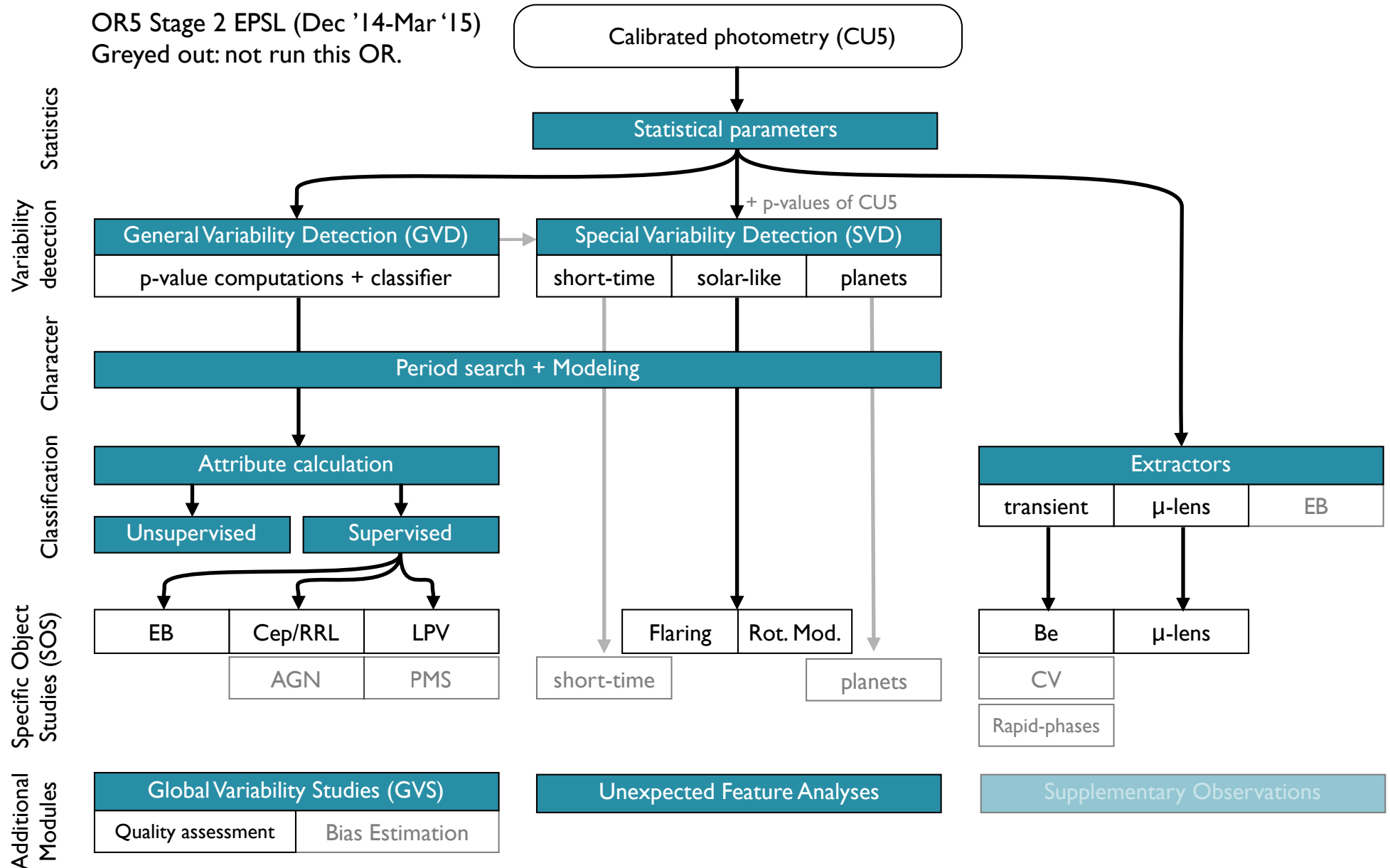




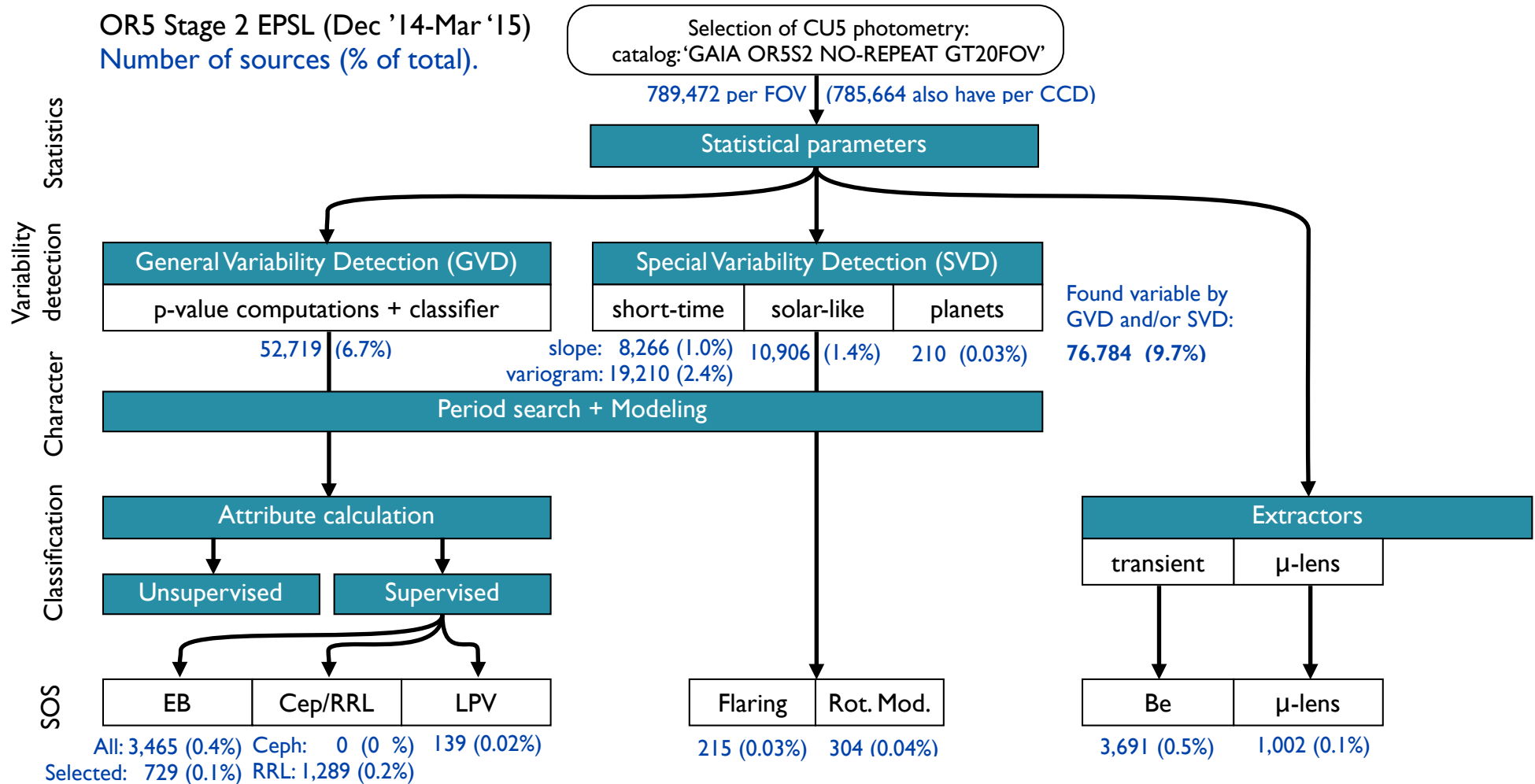
# CU7 processing chain



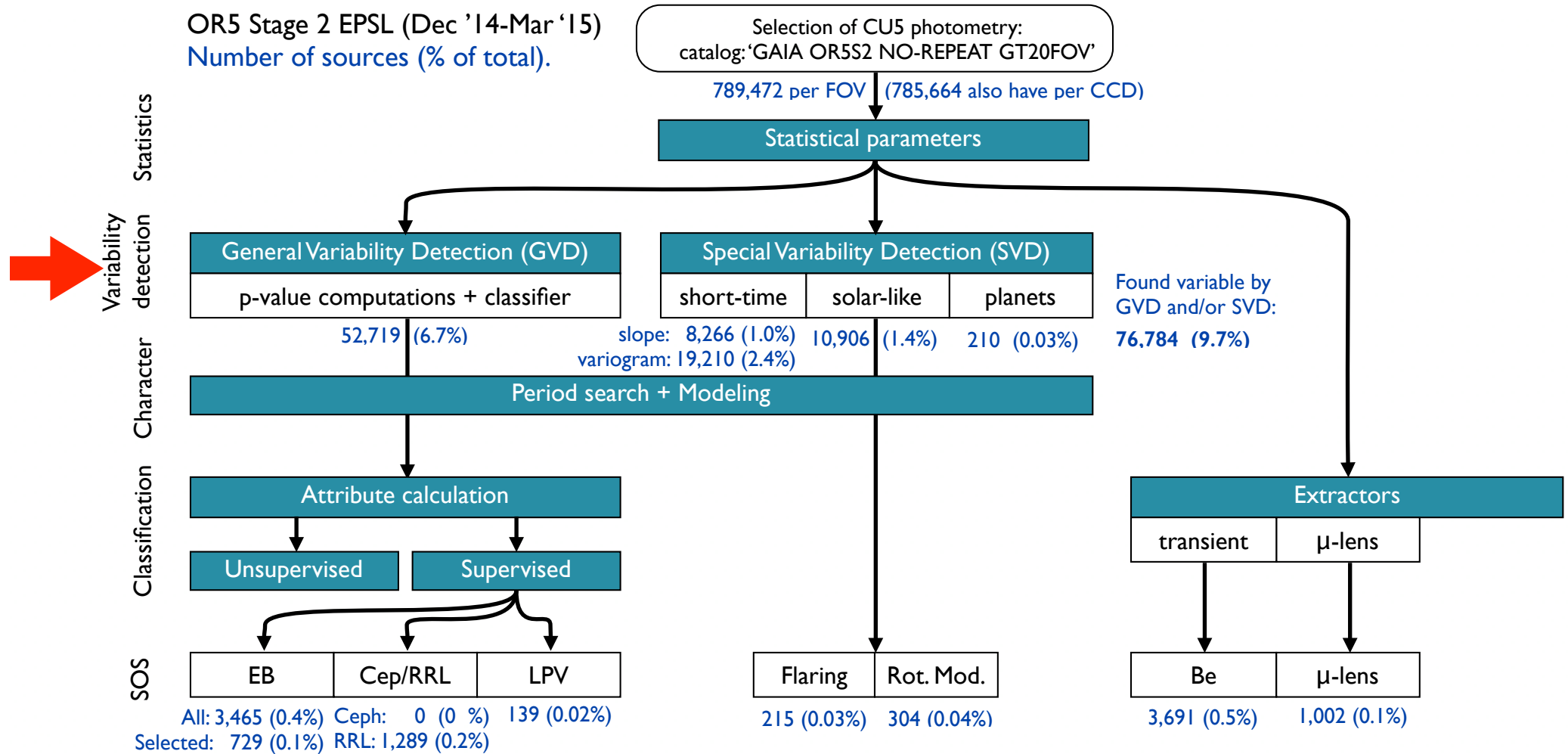
# CU7 processing chain



# CU7 processing chain



# CU7 processing chain



# General Variability Detection

---

Isabelle Lecoeur, Lorenzo Rimoldini, Diego Ordonez, Laurent Eyer

Two fundamental quantities to estimate:

- **Completeness**
- **Contamination**

# General Variability Detection

---

Isabelle Lecoeur, Lorenzo Rimoldini, Diego Ordonez, Laurent Eyer

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)

# General Variability Detection

Isabelle Lecoeur, Lorenzo Rimoldini, Diego Ordonez, Laurent Eyer

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

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

# Special Variability Detection:

---

Elisa Distefano, Shay Zucker, Brandon Tingley, Laurent Eyer, Isabelle Lecoeur, Maroussia Roelens, Leanne Guy, Alessandro Lanzafame

**Statistical Parameters**

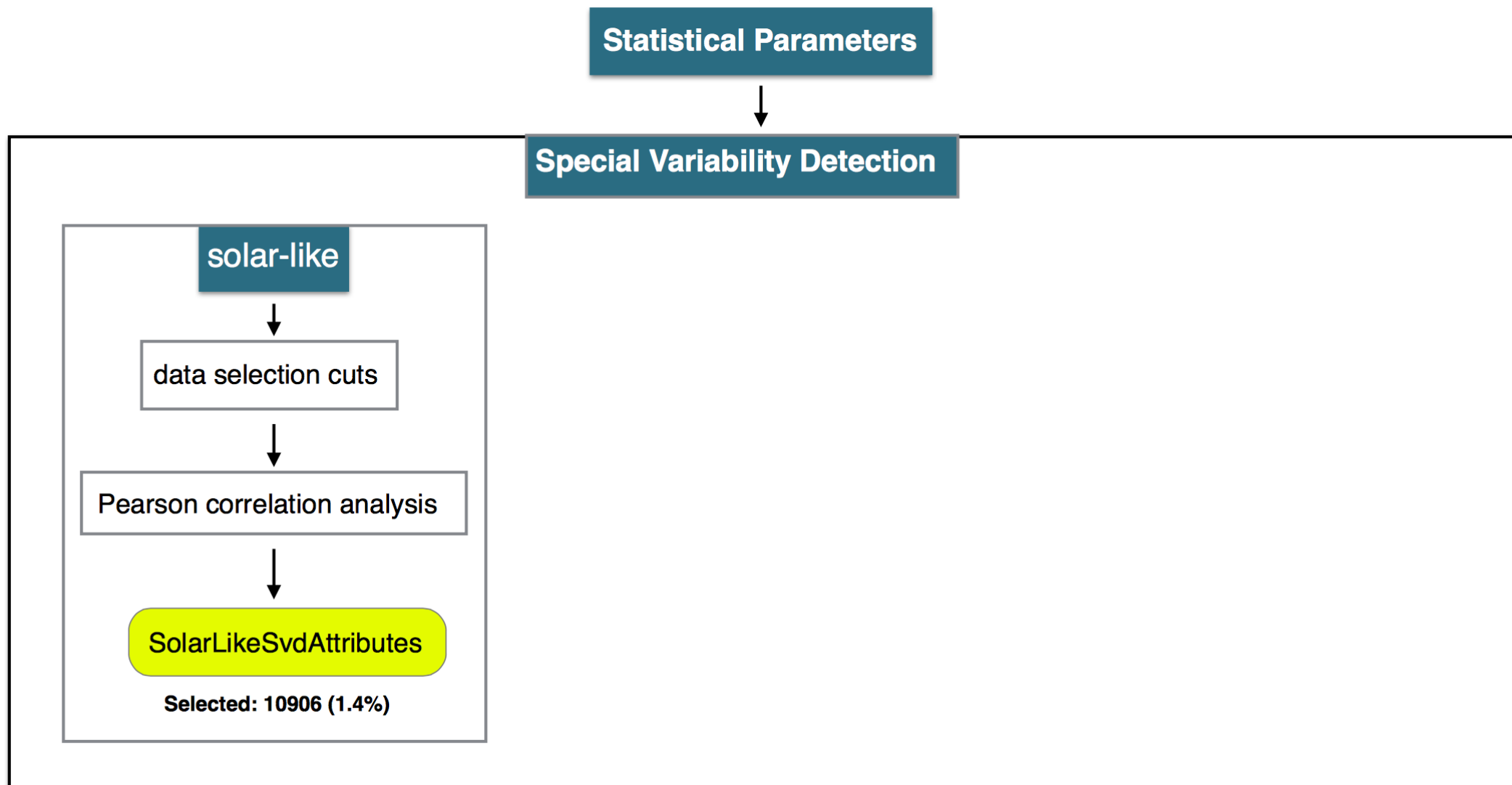


**Special Variability Detection**



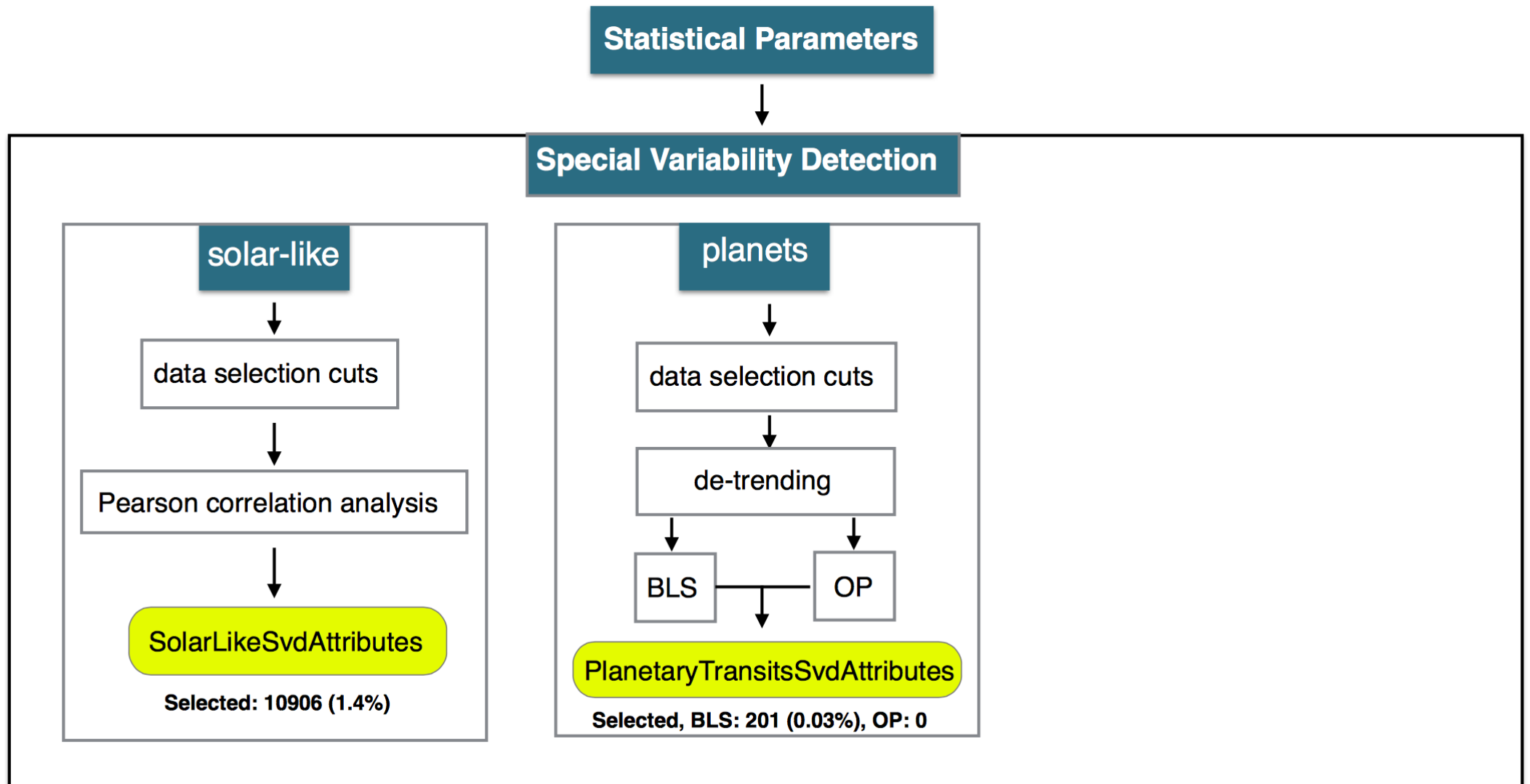
# Special Variability Detection:

Elisa Distefano, Shay Zucker, Brandon Tingley, Laurent Eyer, Isabelle Lecoeur, Maroussia Roelens, Leanne Guy, Alessandro Lanzafame



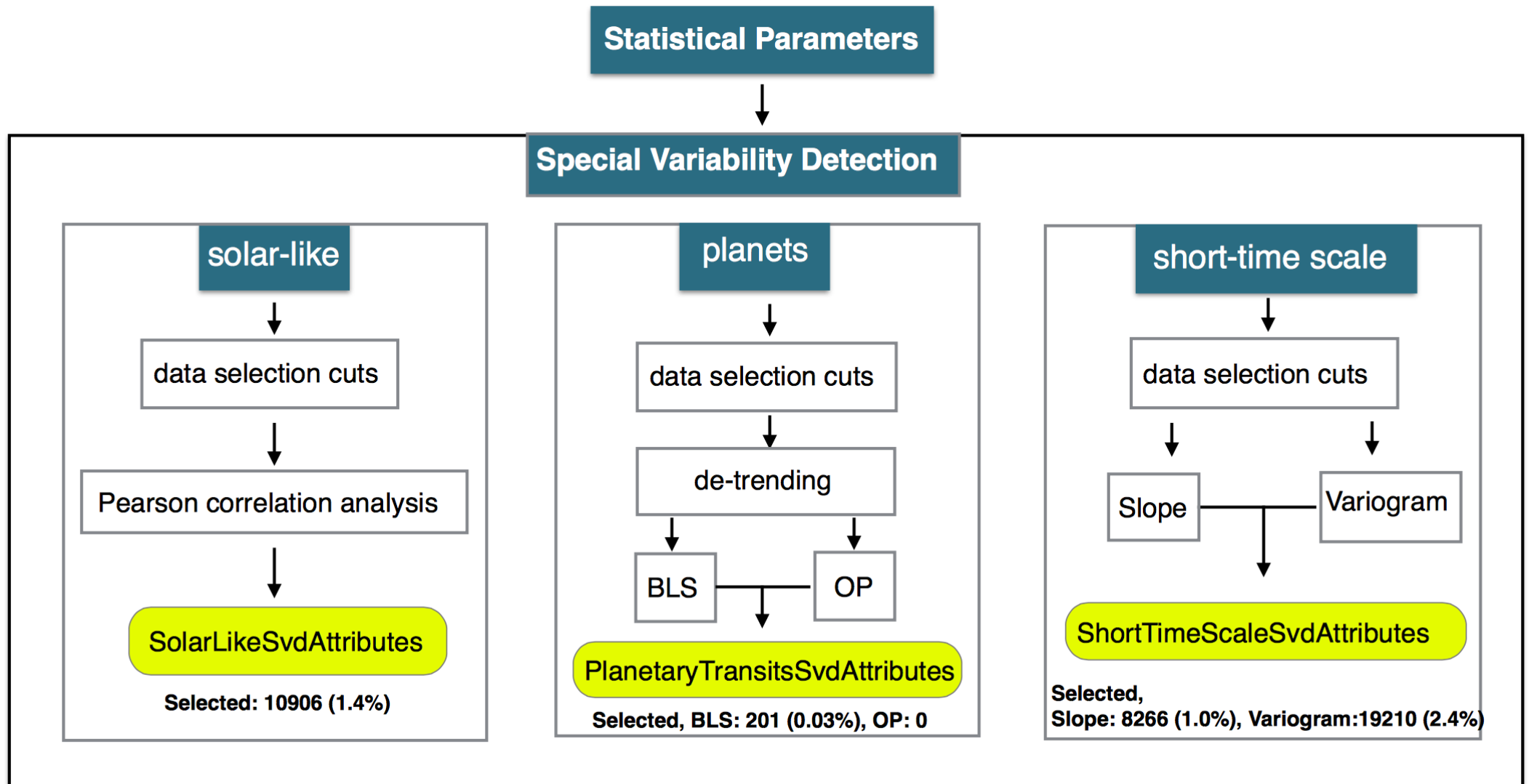
# Special Variability Detection:

Elisa Distefano, Shay Zucker, Brandon Tingley, Laurent Eyer, Isabelle Lecoeur, Maroussia Roelens, Leanne Guy, Alessandro Lanzafame



# Special Variability Detection:

Elisa Distefano, Shay Zucker, Brandon Tingley, Laurent Eyer, Isabelle Lecoeur, Maroussia Roelens, Leanne Guy, Alessandro Lanzafame



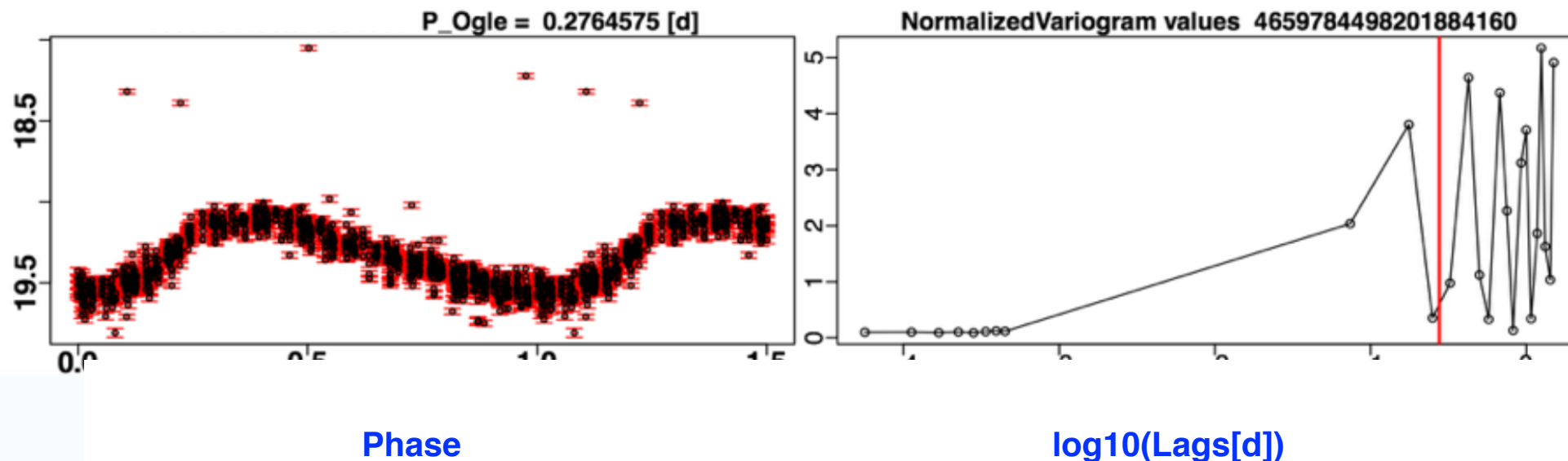
# Special Variability Detection: short time scale

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:

Per-CCD G magnitude



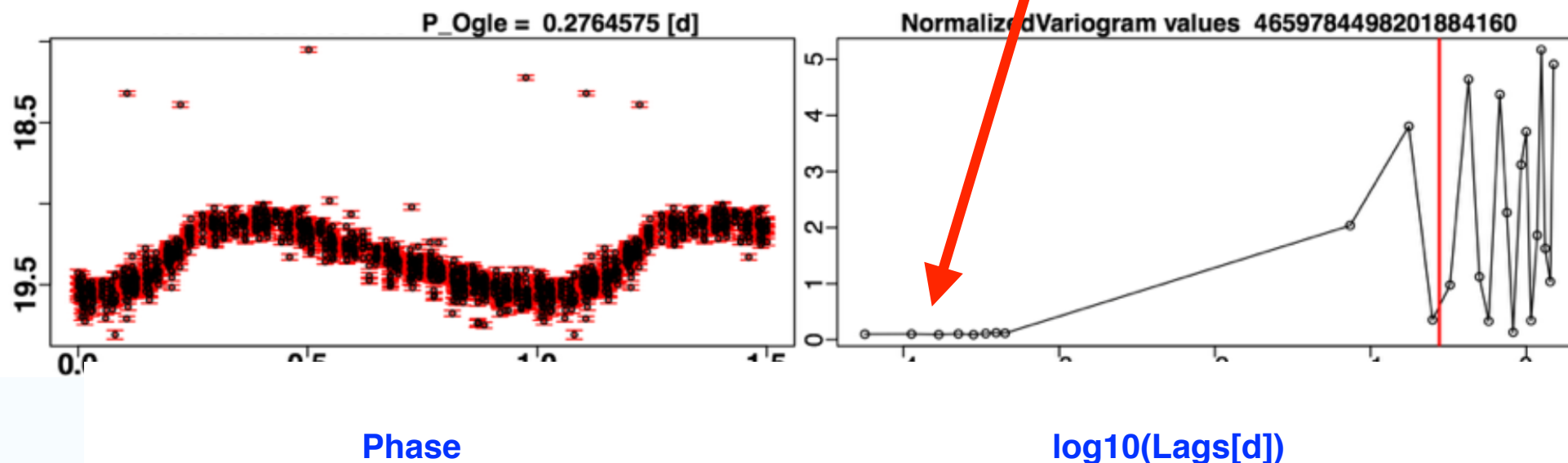
# Special Variability Detection: short time scale

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:

Per-CCD G magnitude



# 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)

# 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)

Box Least Square algorithm gives  
about 200 candidates

# Special Variability Detection: exo-planet transits

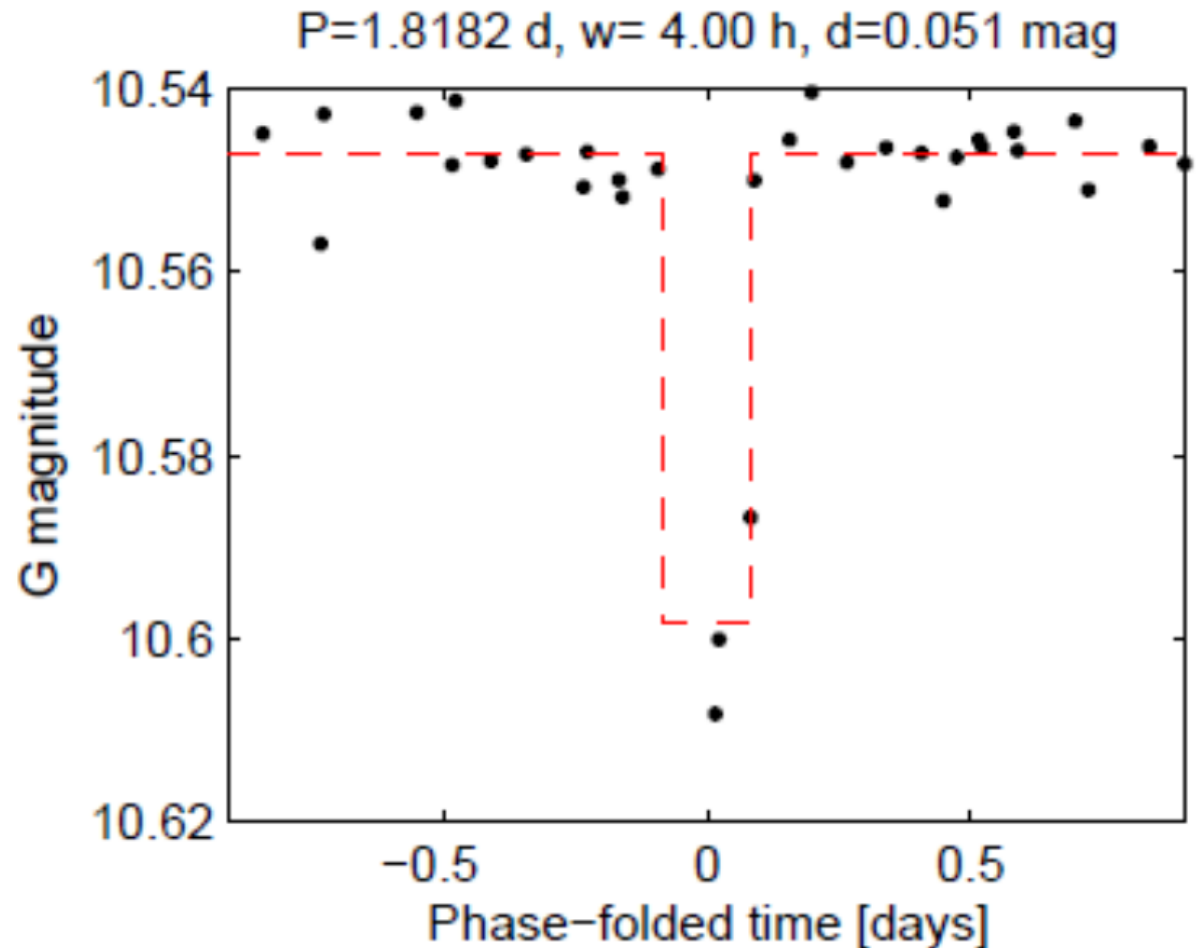
Shay Zucker, Brandon Tingley, Leanne Guy, Alessandro Lanzafame

## Two algorithms:

Box-Least Square

Outlier Probability, Tingley (A&A 2

Box Least Square algorithm gives  
about 200 candidates



Courtesy of S. Zucker



# Special Variability Detection: exo-planet transits

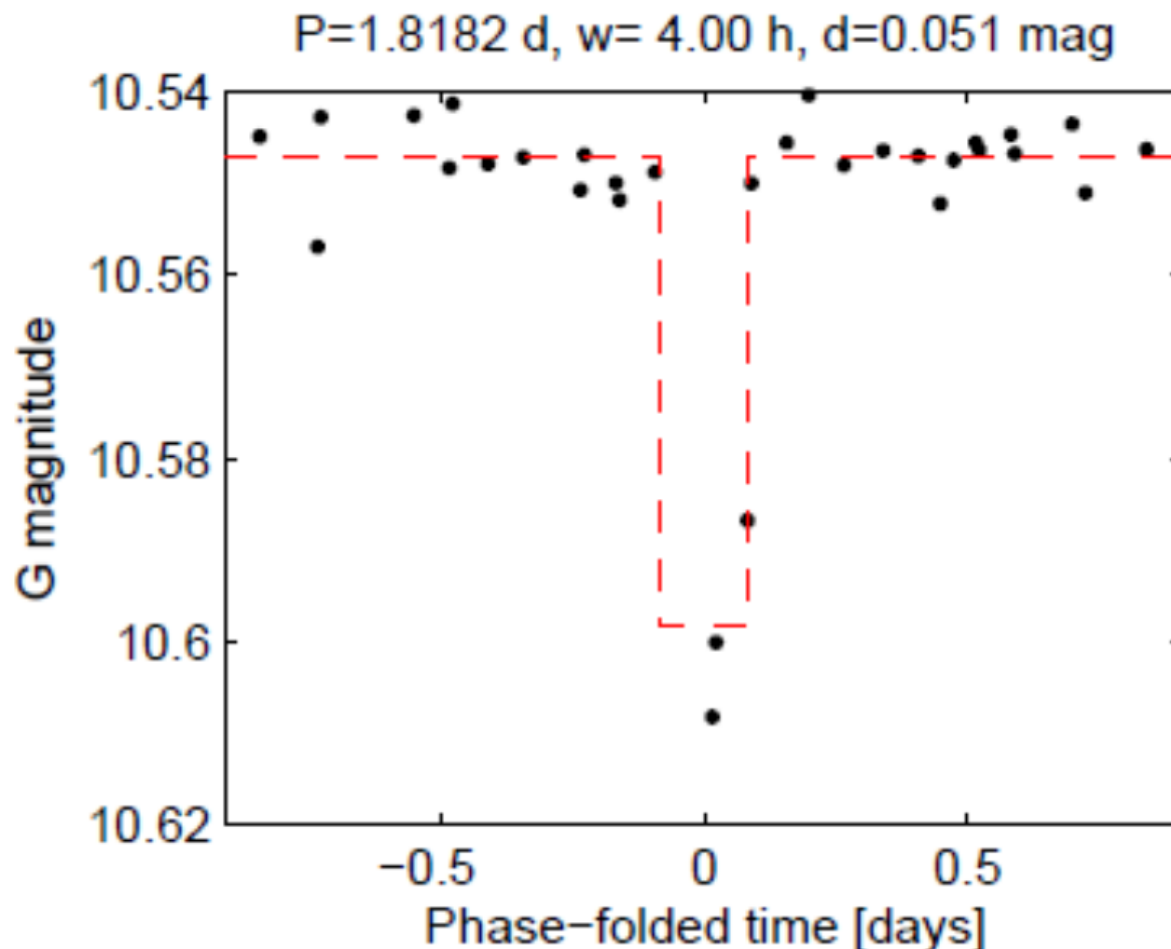
Shay Zucker, Brandon Tingley, Leanne Guy, Alessandro Lanzafame

## Two algorithms:

Box-Least Square

Outlier Probability, Tingley (A&A 2

Box Least Square algorithm gives  
about 200 candidates



Courtesy of S. Zucker

**We do not claim any detection!**  
**Conclusion Box-Least Square is functioning well**

# Special Variability Detection: exo-planet transits

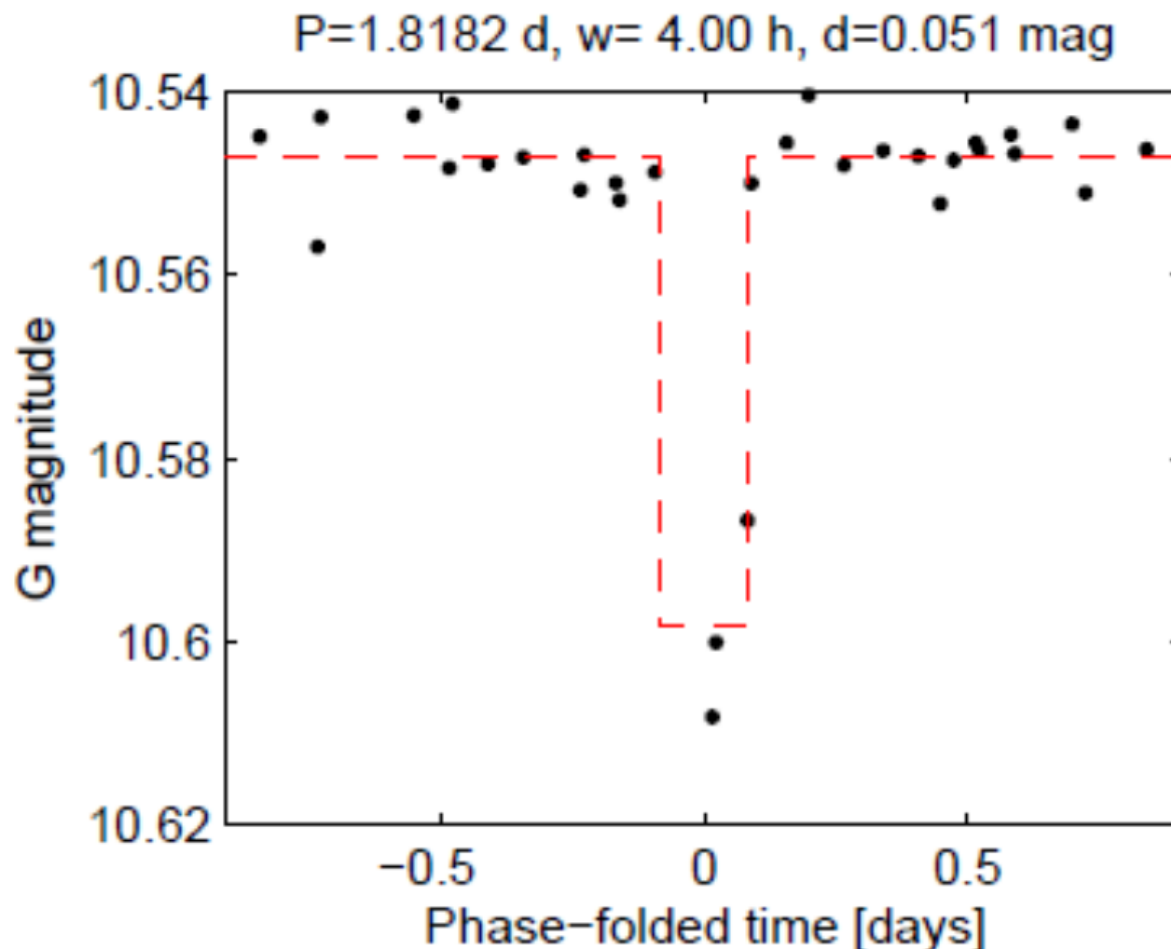
Shay Zucker, Brandon Tingley, Leanne Guy, Alessandro Lanzafame

## Two algorithms:

Box-Least Square

Outlier Probability, Tingley (A&A 2

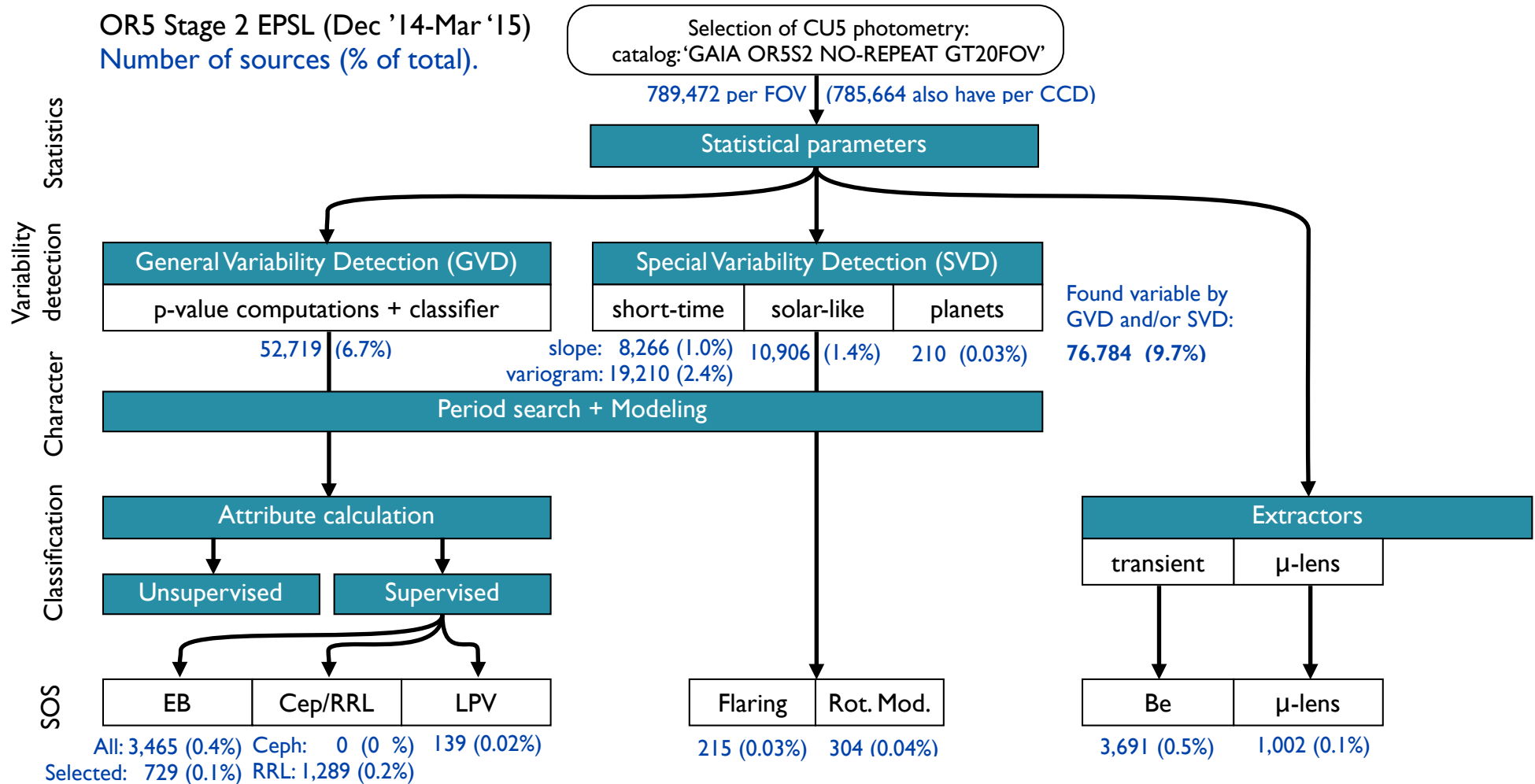
Box Least Square algorithm gives  
about 200 candidates



Courtesy of S. Zucker

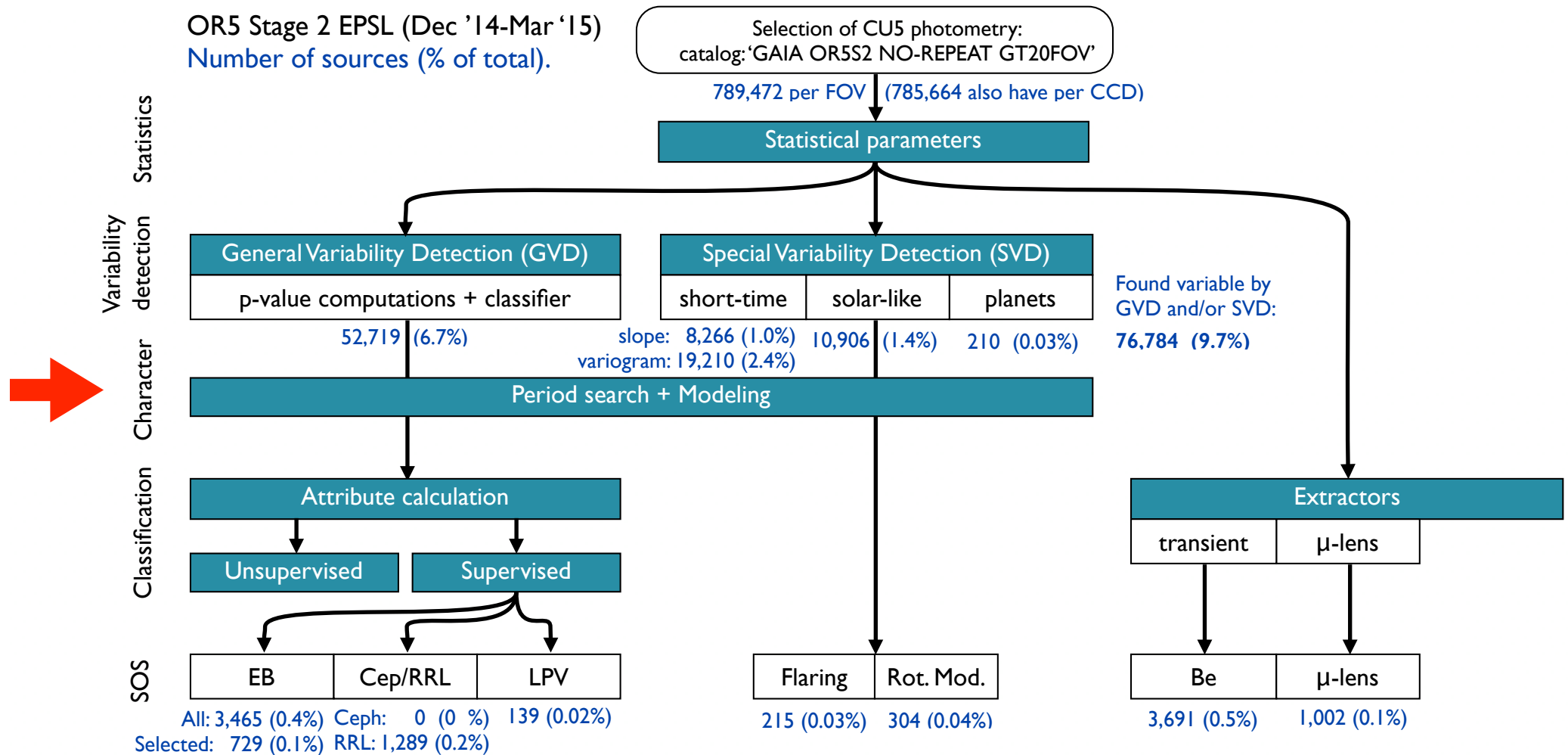
**We do not claim any detection yet !**  
**Conclusion Box-Least Square is functioning well**

# CU7 processing chain



# CU7 processing chain

OR5 Stage 2 EPSL (Dec '14-Mar '15)  
 Number of sources (% of total).



# Characterisation

---

Jan Cuypers, Leanne Guy, Lorenzo Rimoldini, Joris De Ridder

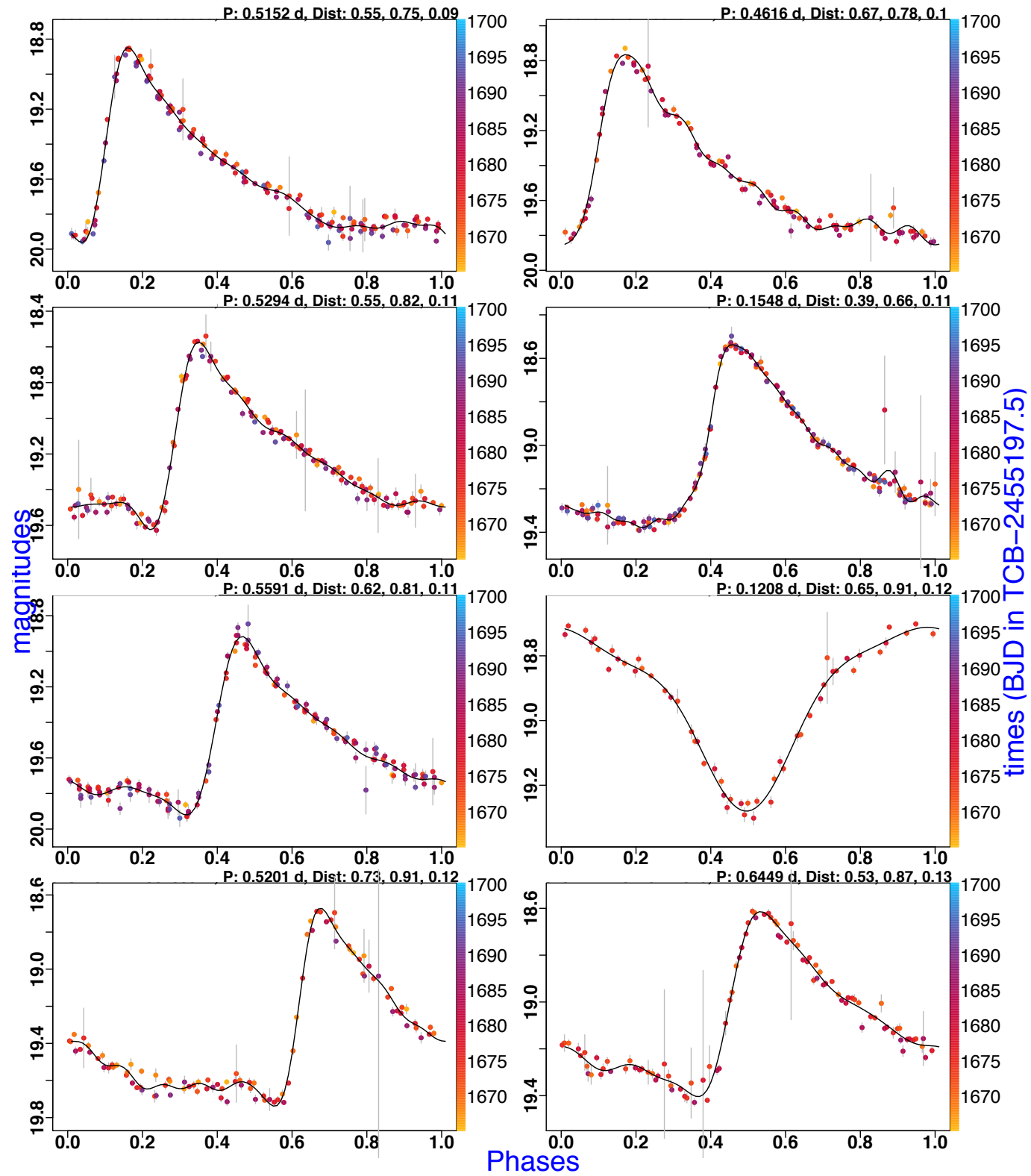
Time series per object:

Time(i), G-, BP-, RP- mag(i) [ or radial velocity(i) ]  $i=1, \dots$ , number of measurements

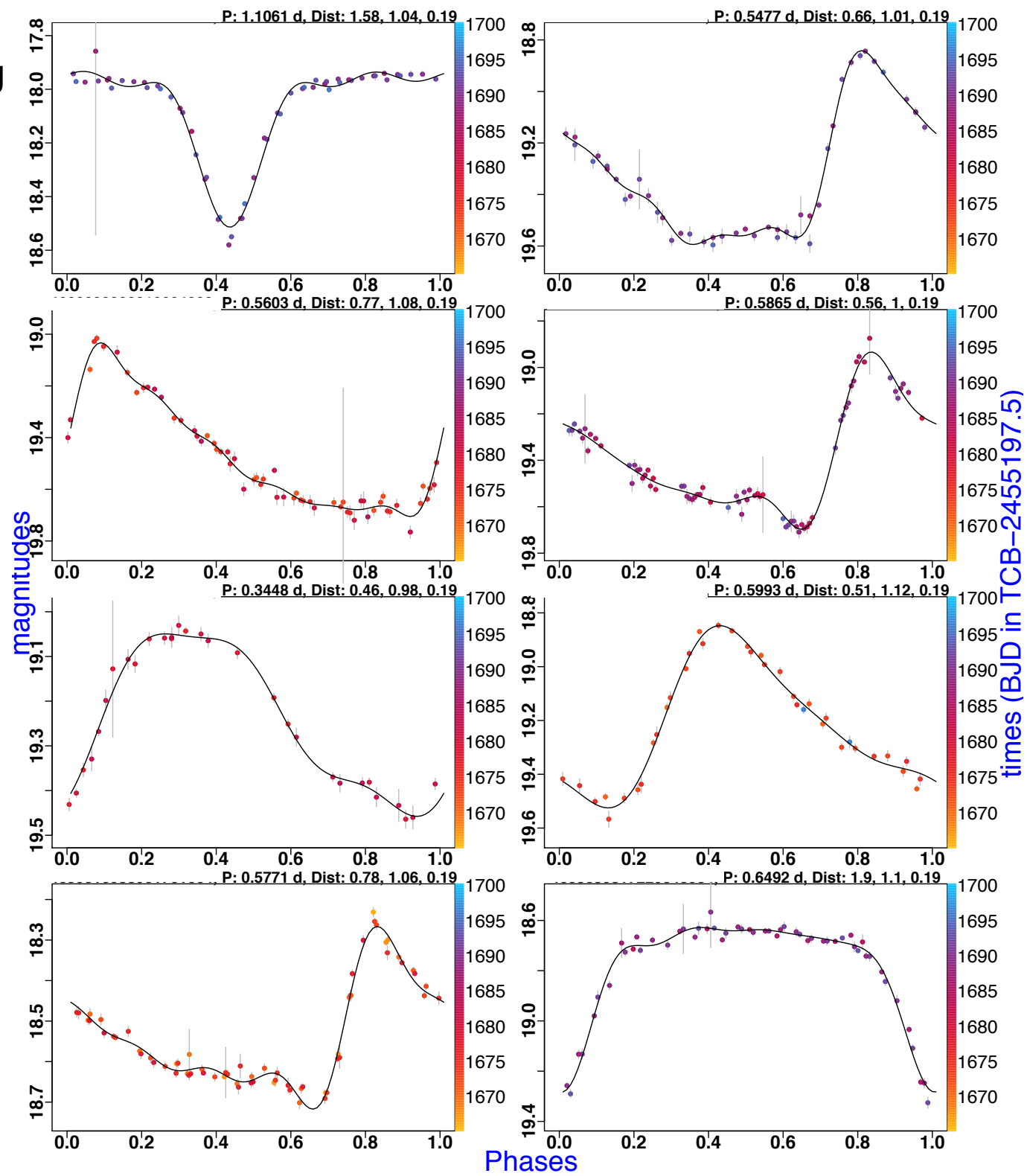
Goal: To define attributes

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

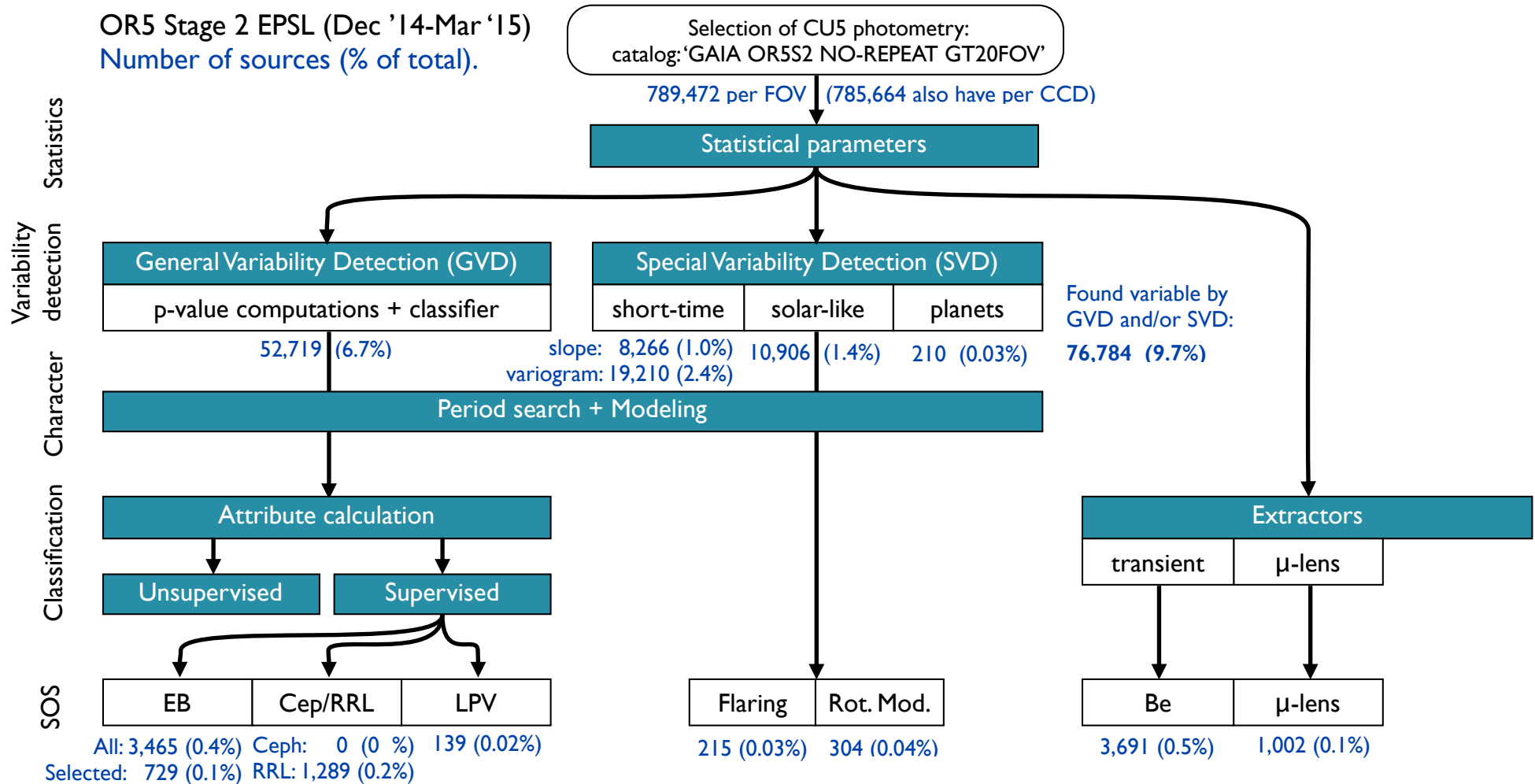
# Few examples of modelling



# Few examples of modelling



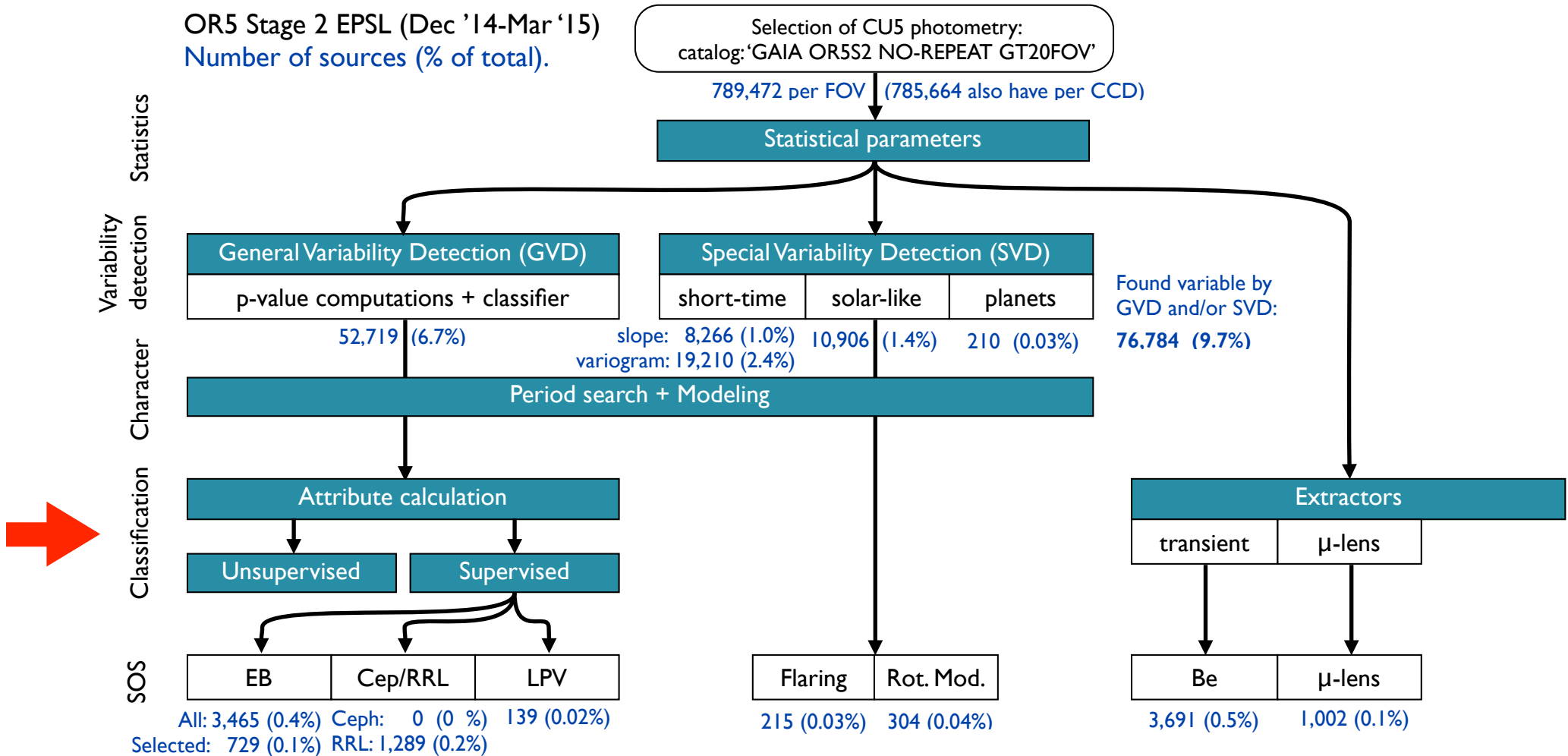
# CU7 processing chain





# CU7 processing chain

OR5 Stage 2 EPSL (Dec '14-Mar '15)  
 Number of sources (% of total).



# Classification

---

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

**Supervised classification (several methods):**

# 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

# Classification

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

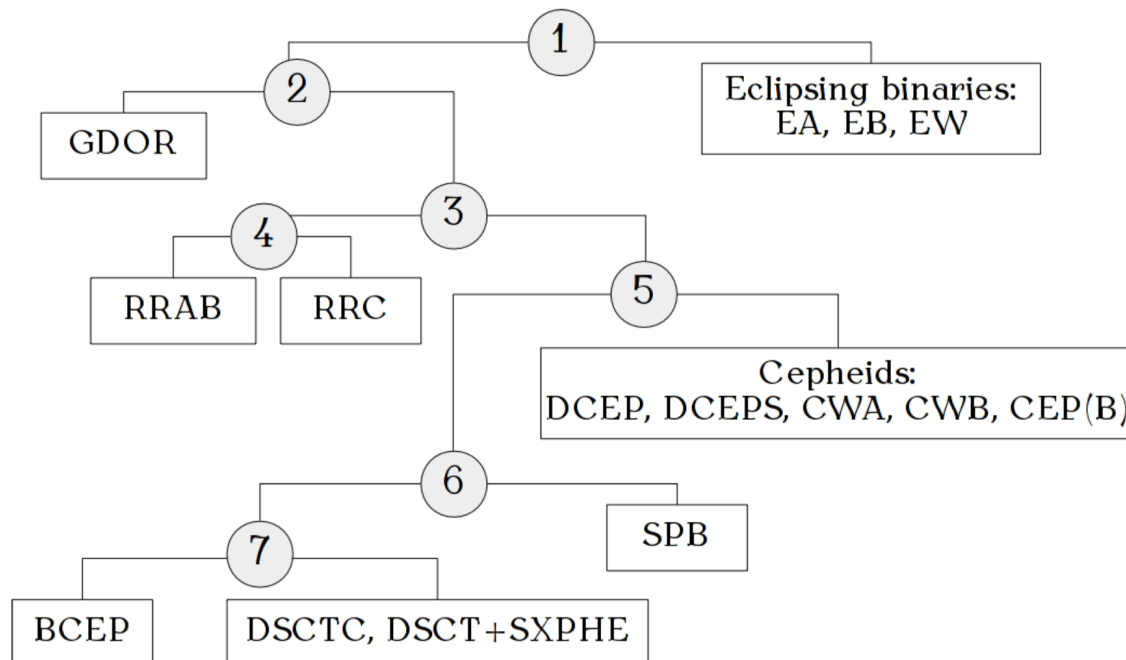
## Supervised classification (several methods):

Multistage tree:  
Bayesian networks

Multistage tree:  
Gaussian mixture

Random Forest

Tree for Gaussian Mixture:



# Classification

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

## Supervised classification (several methods):

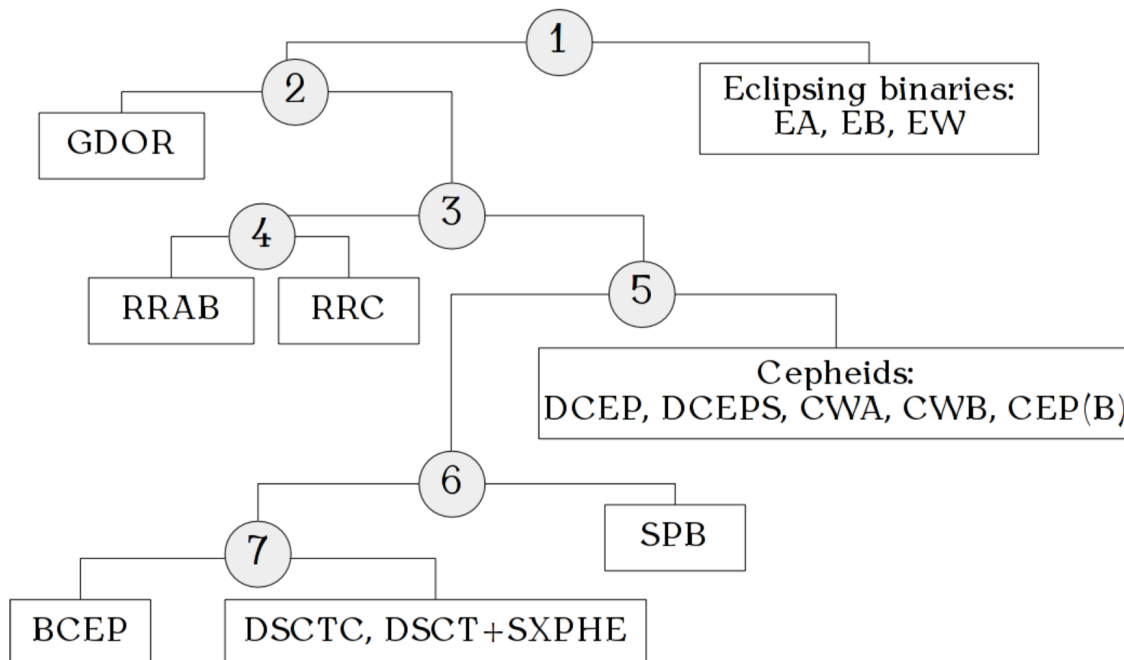
Multistage tree:  
Bayesian networks

Multistage tree:  
Gaussian mixture

Random Forest

Furnish training set  
built from Crossmatched data

Tree for Gaussian Mixture:



# Classification

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

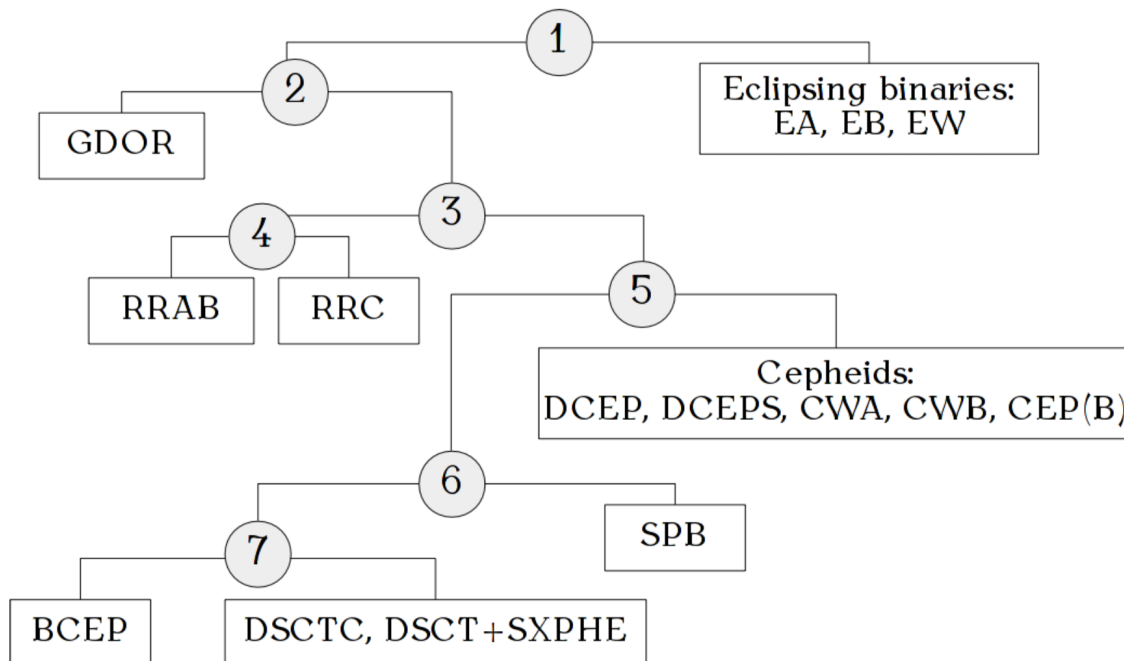
## Supervised classification (several methods):

Multistage tree:  
Bayesian networks

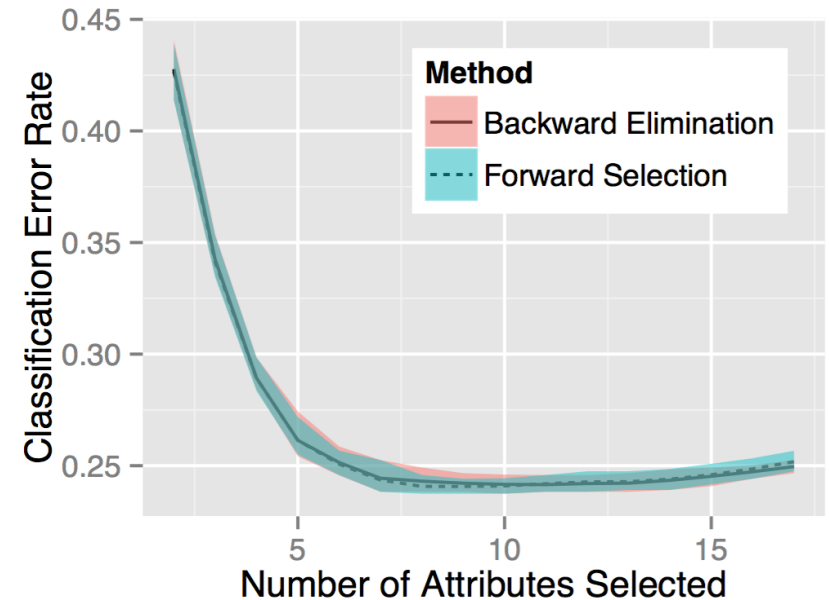
Multistage tree:  
Gaussian mixture

Random Forest

Tree for Gaussian Mixture:



Furnish training set  
built from Crossmatched data



# Classification: Attributes for Random Forest

(a) Attribute Ranking

SKEWNESS_PERCENTILE_5	60	5		30		1	5	
LOG_QSO_VAR		95		3			2	
LOG_QSO_PROB			44	14	14			
BP_MINUS_RP_COLOR	27		7	28	24		1	
LOG_PERIOD	11		2	24	59	1	1	
LOG_AMPLITUDE	1		1		1	78	2	1
MODEL_POLYCOEFF_1STDEGREE			2		1	16	62	
LOG_SCATTER_RES_RAW	1				1	3	16	16
	1	2	3	4	5	6	7	8

Attribute Rank

(b) Attribute Importance

	CONSTANT	QSO	ECL	OTHER	RRLYR	LPV	ELL	DSCT	DCEP	UG
SKEWNESS_PERCENTILE_5			Dark		Light					
LOG_QSO_VAR	Light	Dark	Dark	Light	Black					
LOG_QSO_PROB		Dark								
BP_MINUS_RP_COLOR		Light		Light	Light					
LOG_PERIOD										
LOG_AMPLITUDE	Light				Light					
MODEL_POLYCOEFF_1STDEGREE	Light									
LOG_SCATTER_RES_RAW	Light	Light			Light					

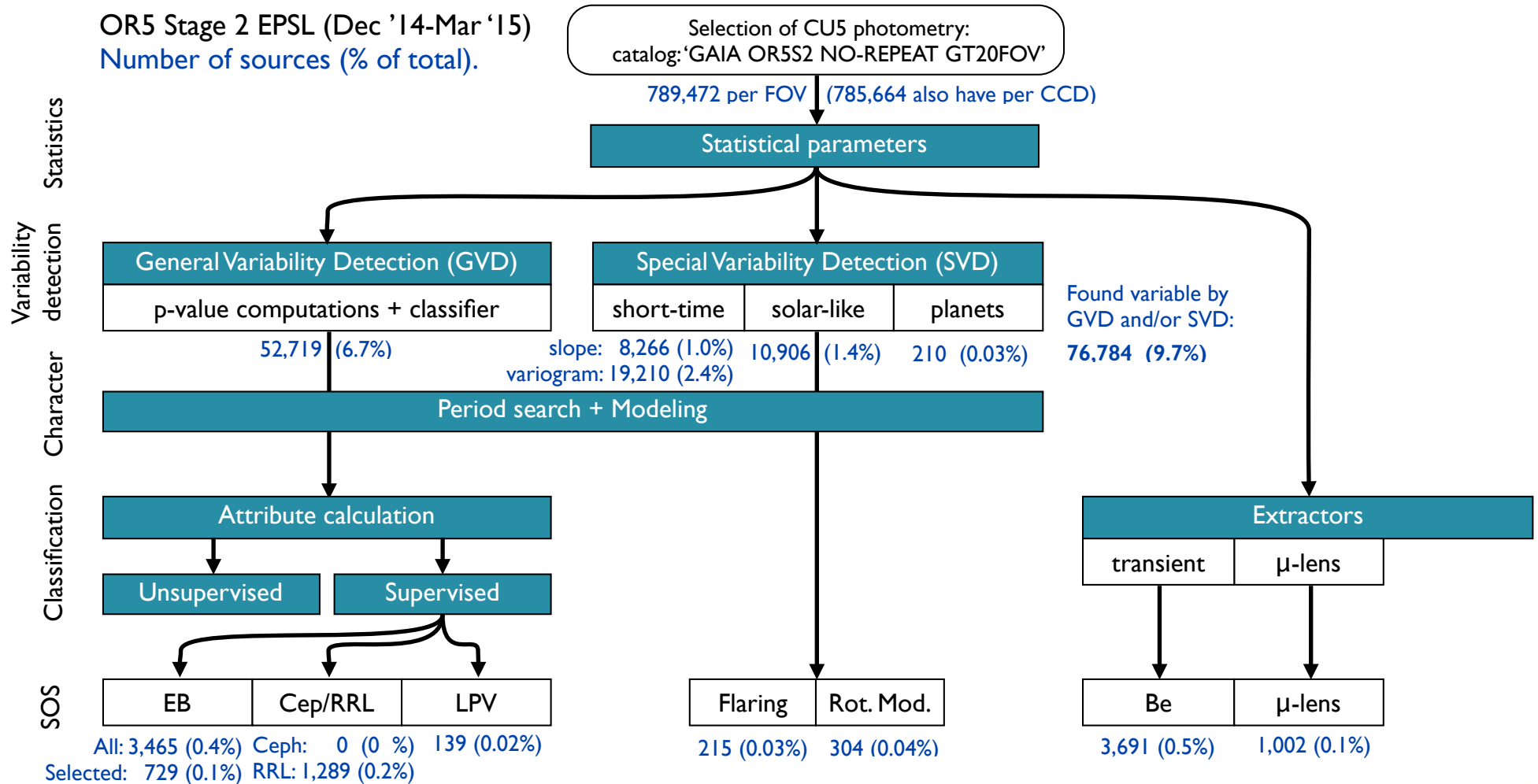
# Classification

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				
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	-	-	-

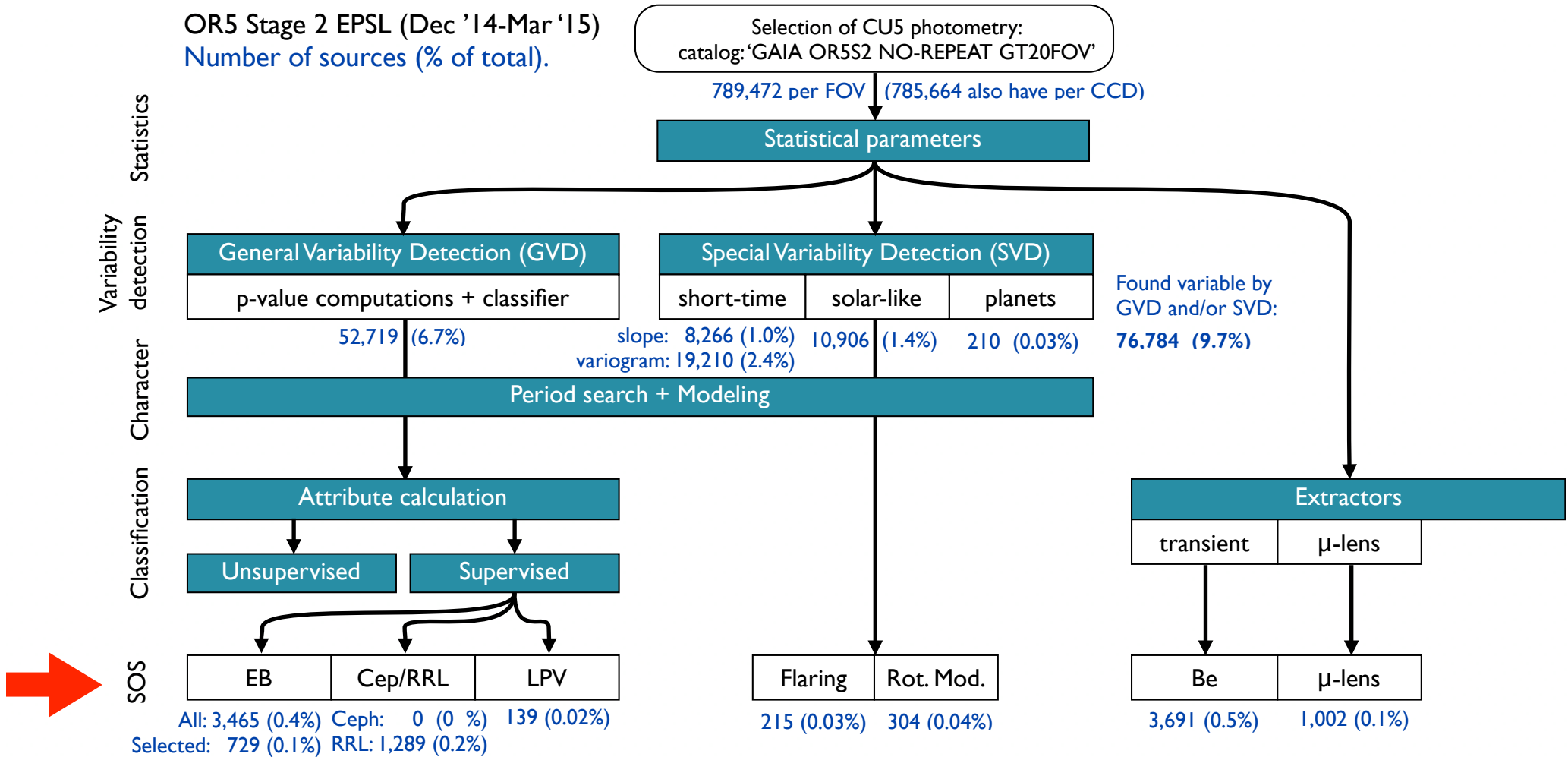


# CU7 processing chain



# CU7 processing chain

OR5 Stage 2 EPSL (Dec '14-Mar '15)  
 Number of sources (% of total).



# Specific Object Studies

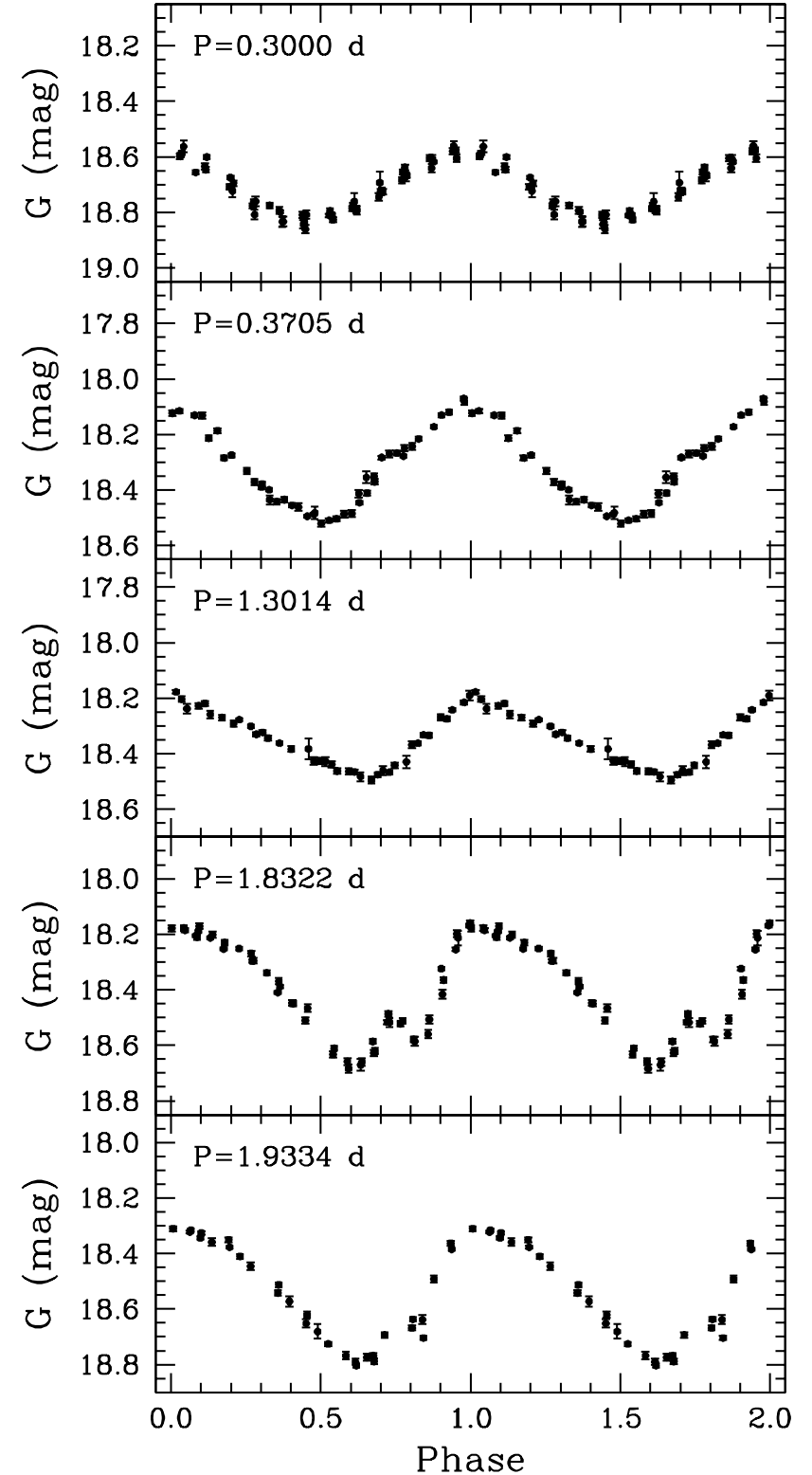
## RR Lyrae and Cepheid stars

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

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 Lecoer-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

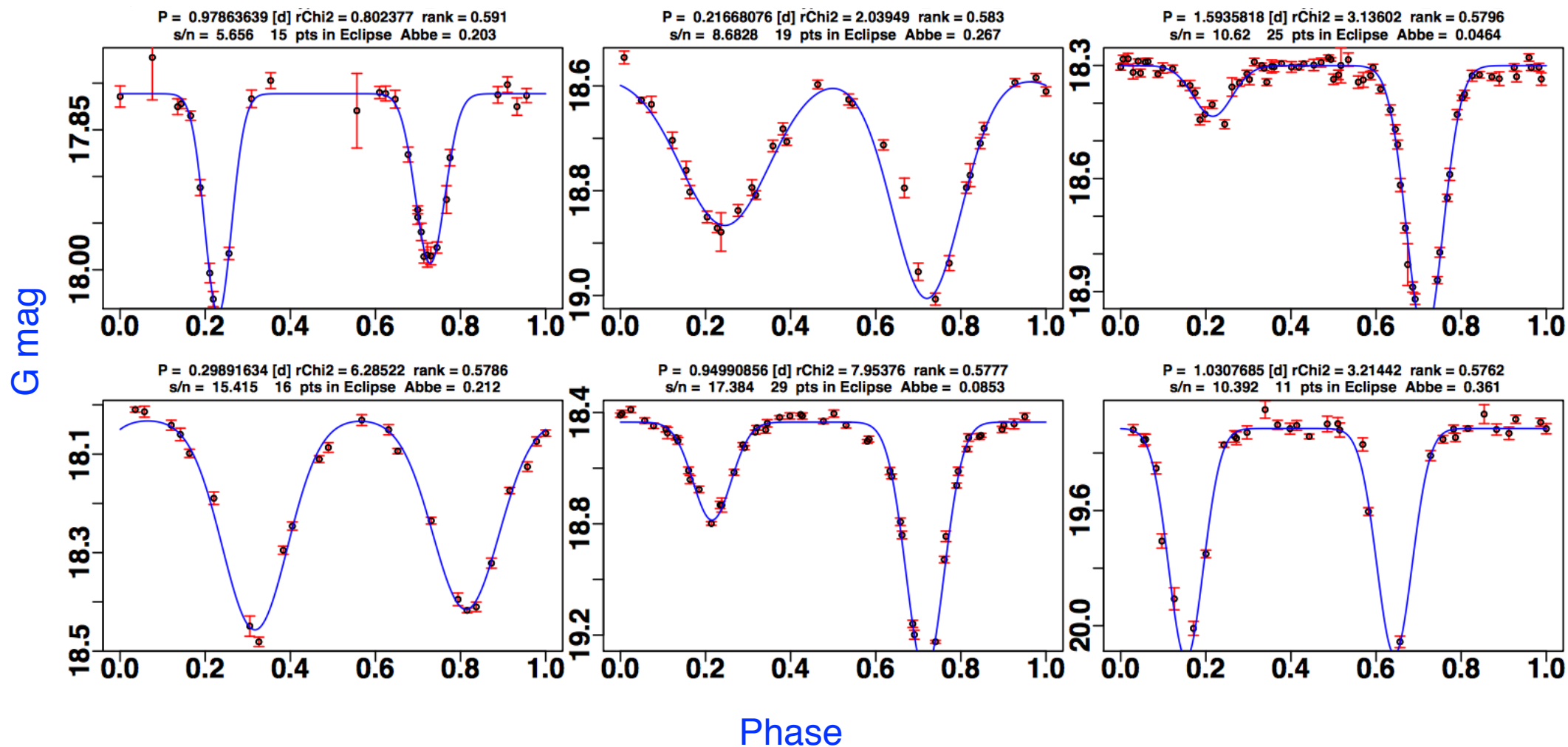
Nami Mowlavi, Berry Holl, Isabelle Lecoeur, Fabio Barblan, Lorenzo Rimoldini

Eclipsing binaries go to CU4 for a full modelling

In CU7, some simple modelling are made

Solutions allow a ranking

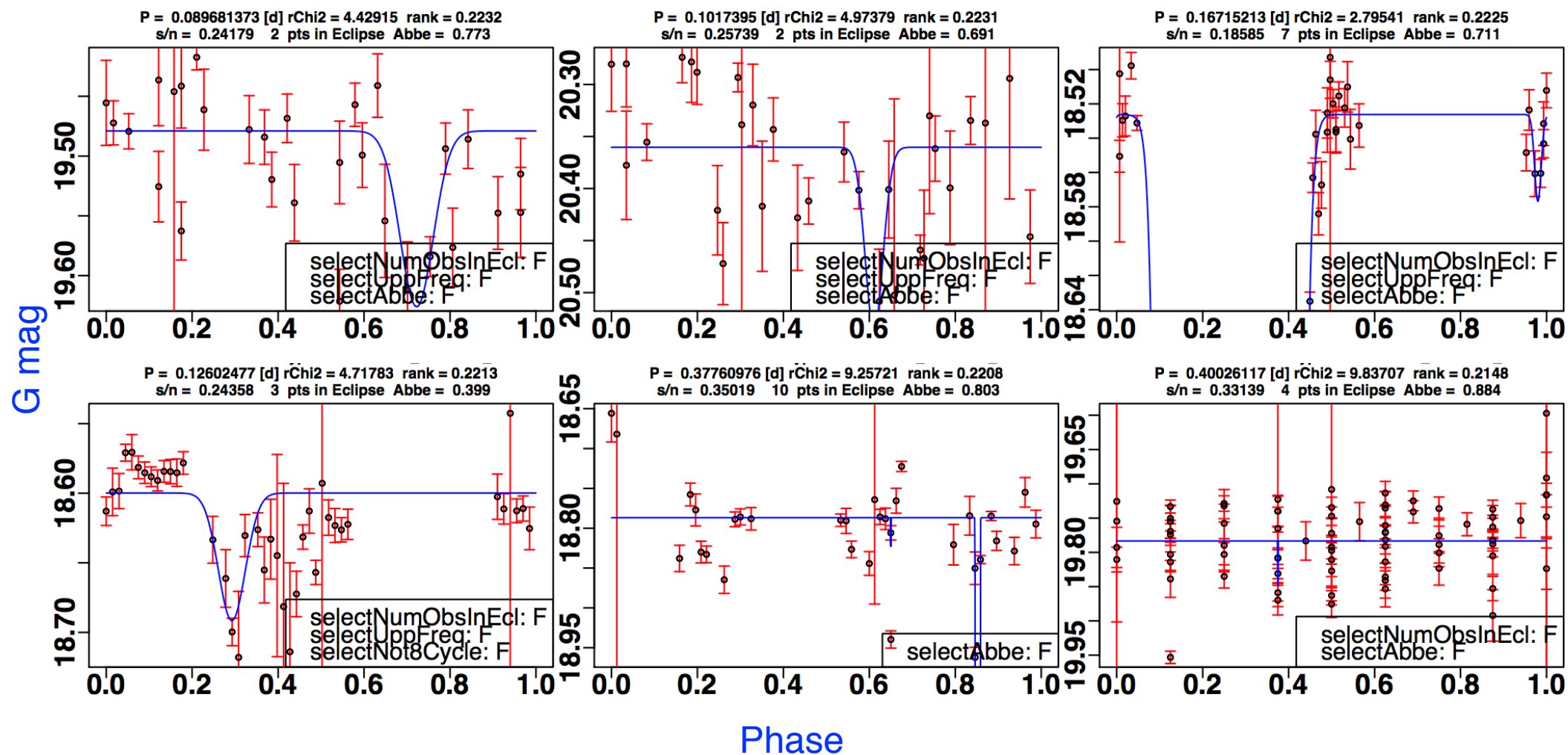
Highest rank



# Specific Object Studies: Eclipsing binaries

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

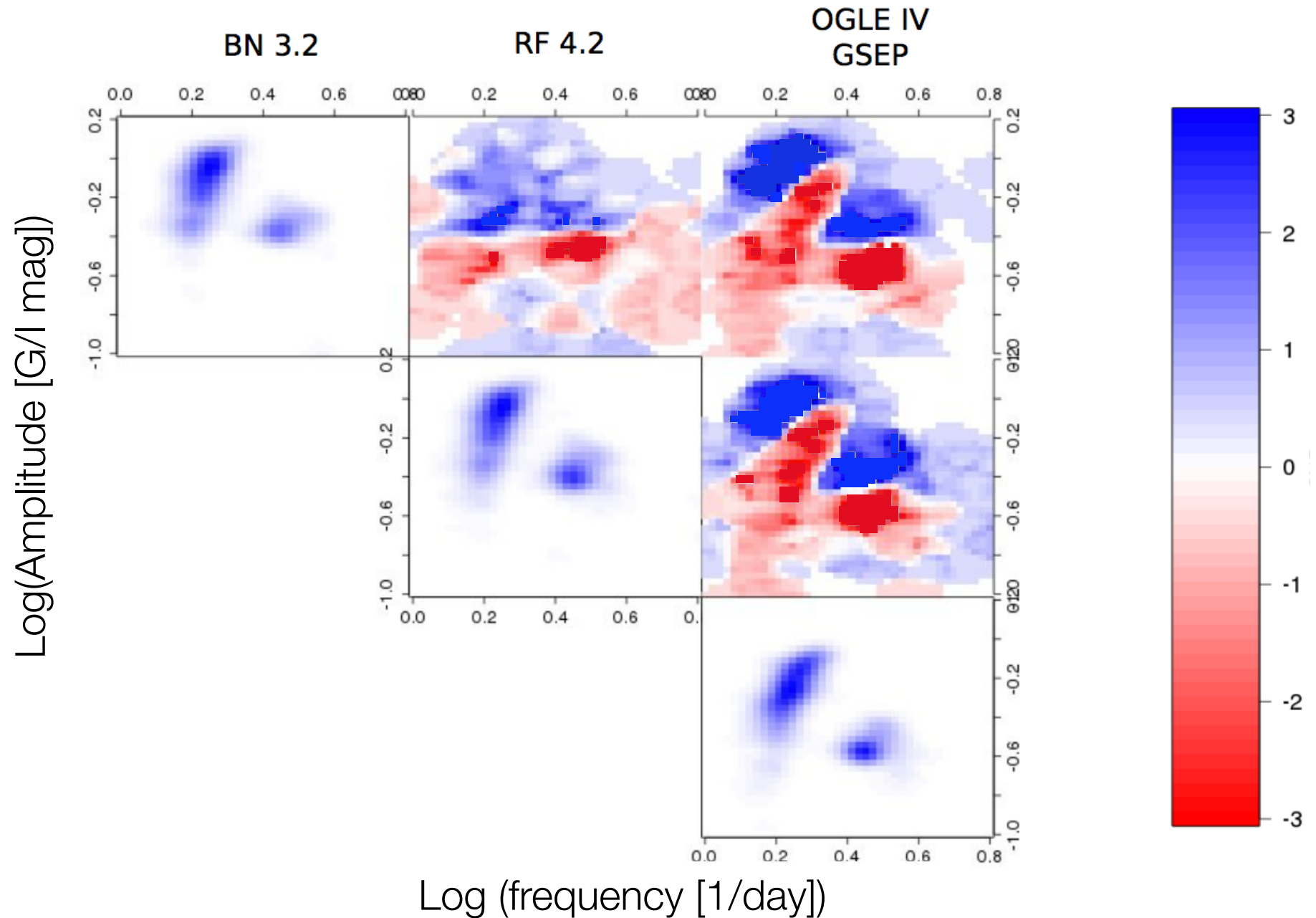
Lowest rank



# Global Variability studies

Luis Sarro, Manuel, André Moitinho, Leanne Guy

## Comparison of distribution functions of RR Lyrae stars



# Conclusions

---

# Conclusions

---

- We can remark that for a first reduction, the quality of G band is remarkable
- The photometry will be improved
- The Variability Processing and Analysis seem to be on the right track !
- We may release the EPSL data variability analysis (Data Release 1 or 2)