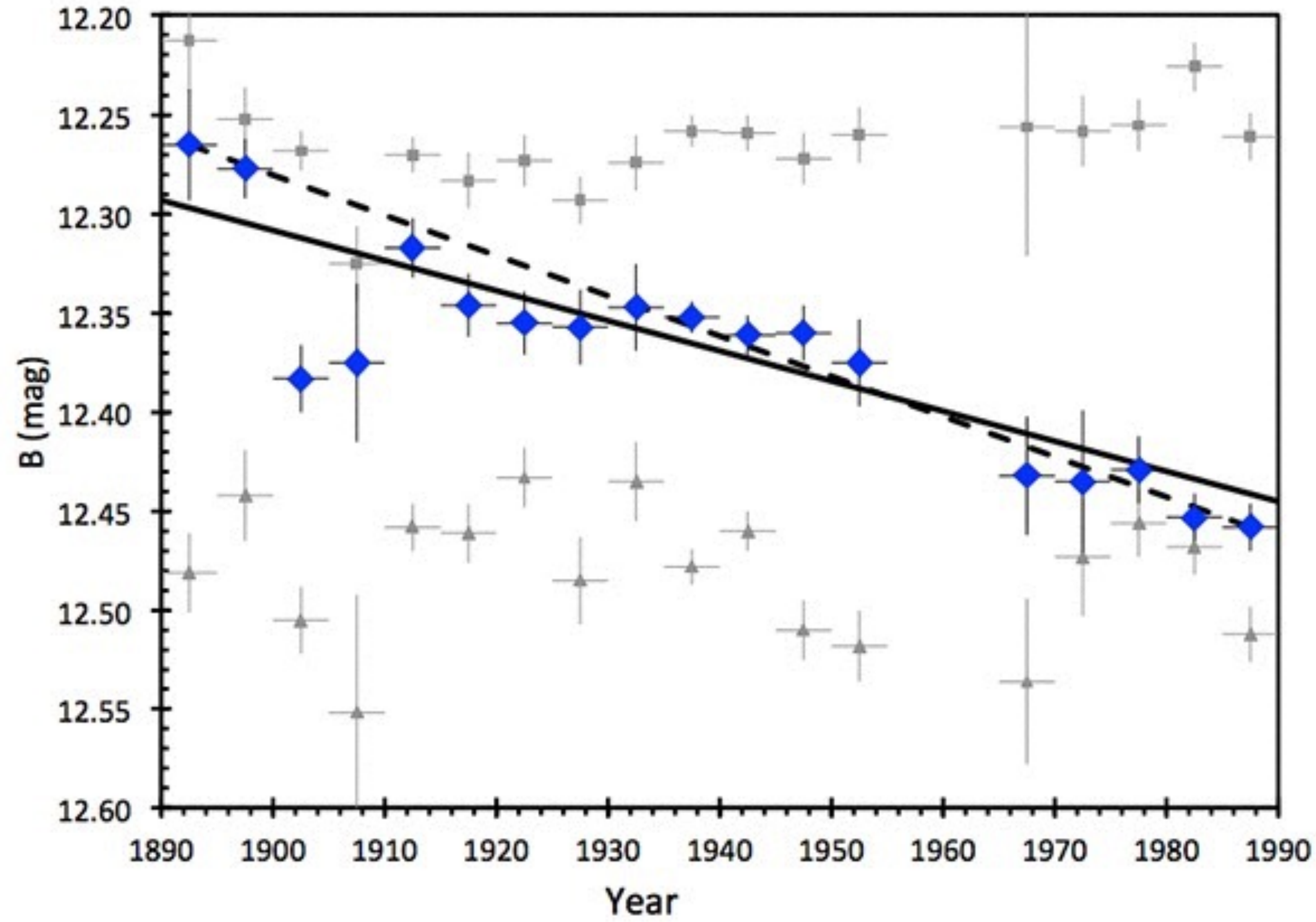
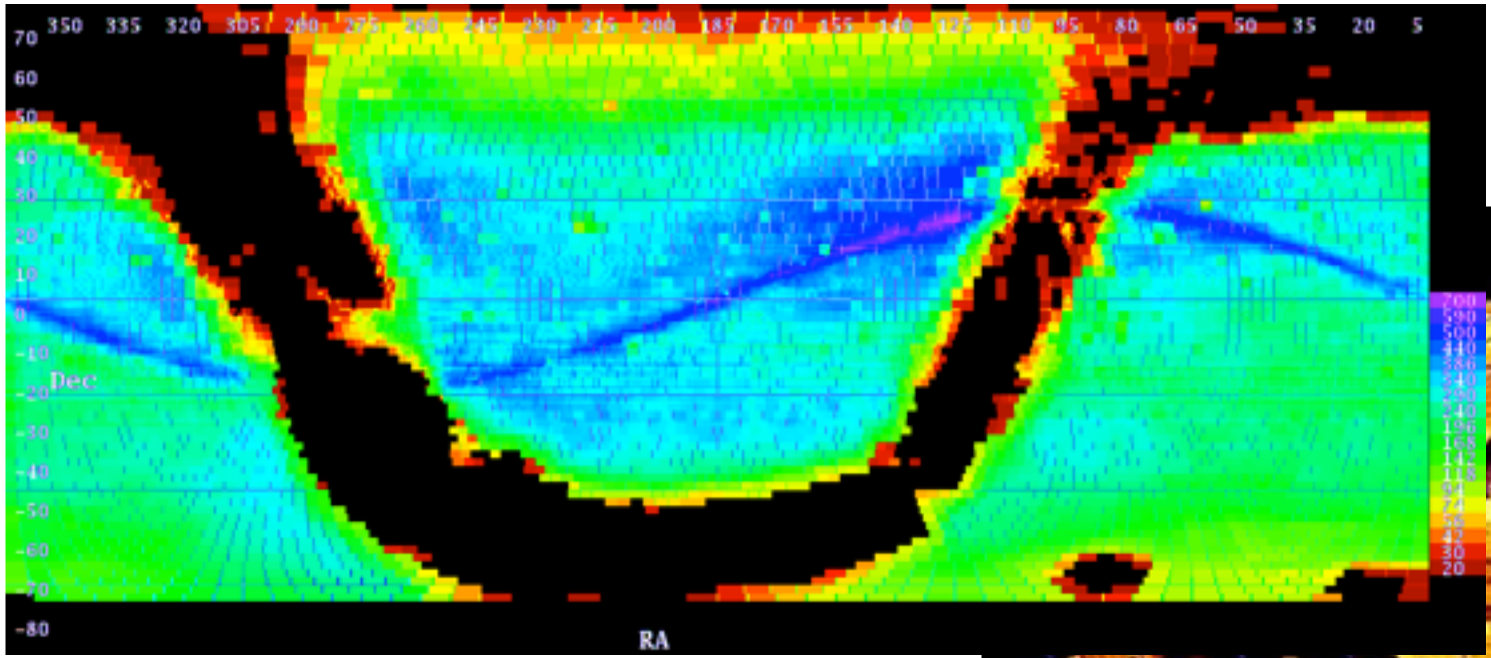
Complete Classification Conundrum



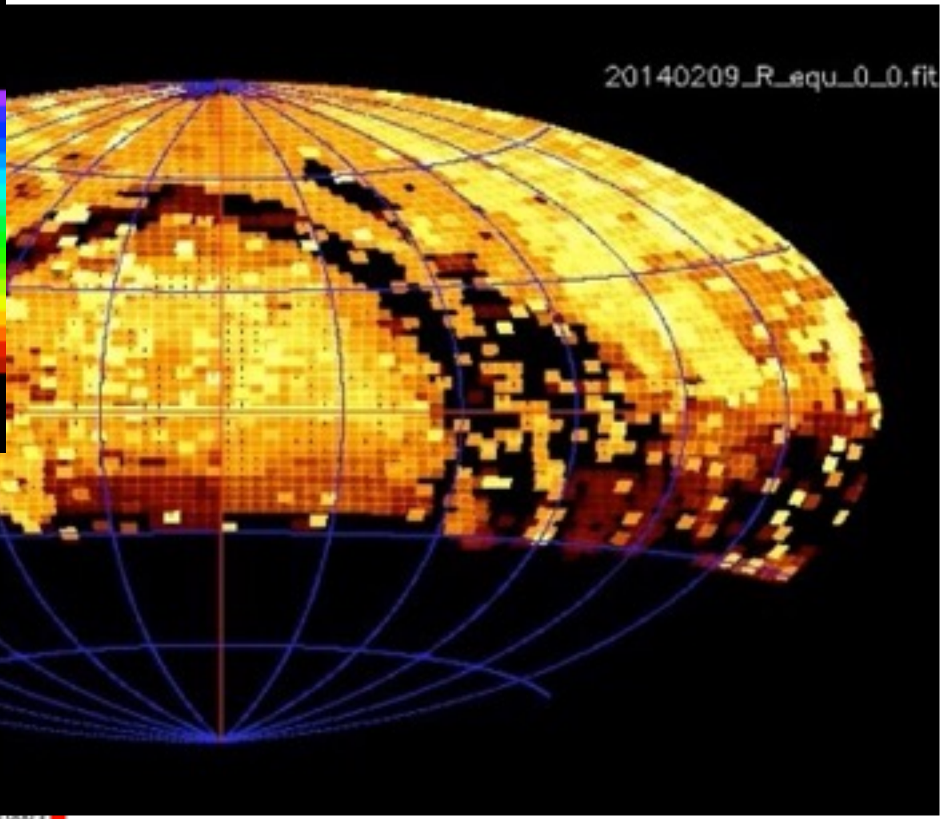
Ashish Mahabal
aam@astro.caltech.edu
Center for Data-Driven Discovery (CD^3), Caltech
Collaborators: CRTS, PTF, LSST, SAMSI, IUCAA … teams
SCMA VI, CMU, 20160607

Sky Maps of a few (optical) surveys

CRTS

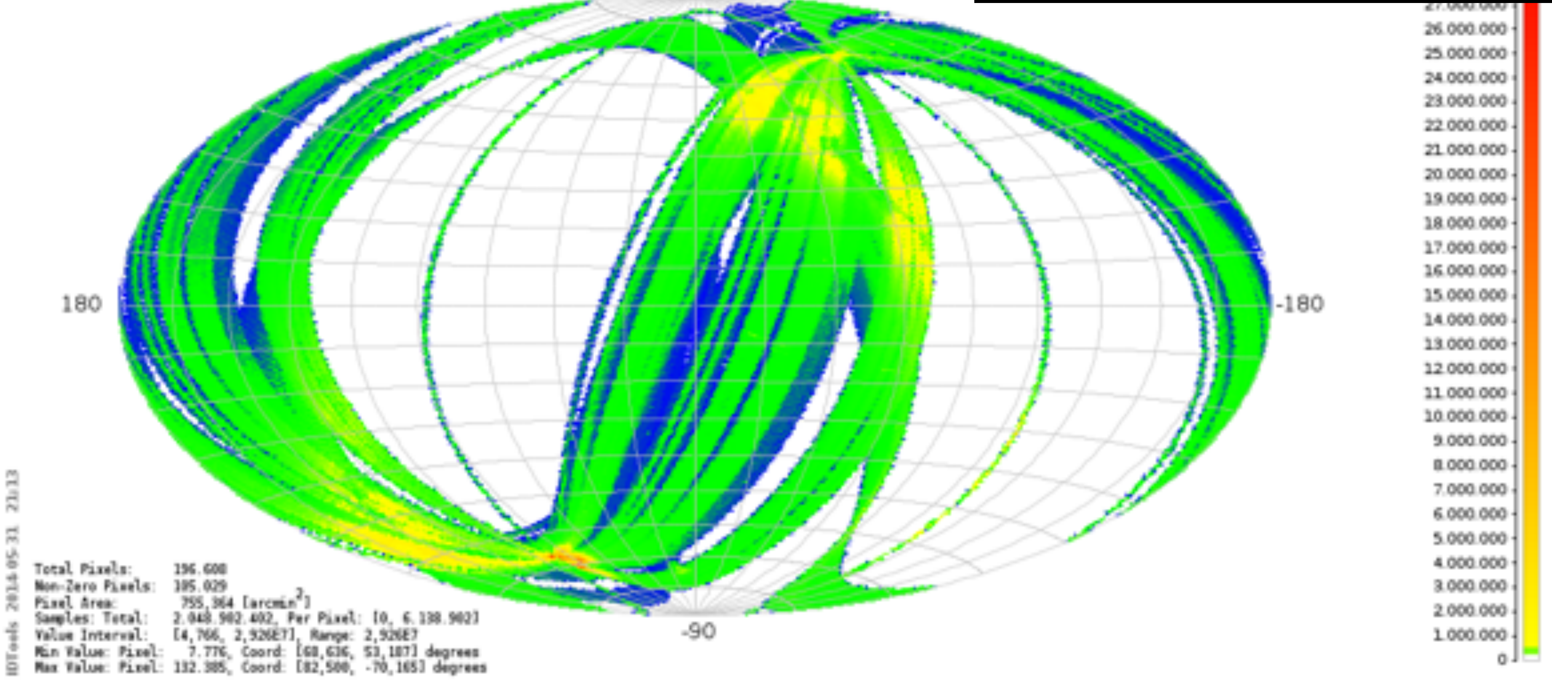


PTF

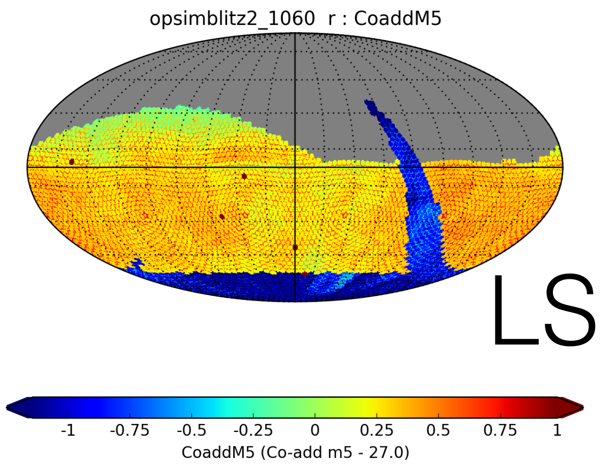


Gaia

Observed sky [obs/deg²]



```
Total Pixels:       196.608
Non-Zero Pixels:    105.029
Pixel Area:          755,364 [arcmin²]
Samples: Total:     2.048.902.402, Per Pixel: [0, 6.138.902]
Value Interval:     [4,766, 2,926E7], Range: 2,926E7
Min Value: Pixel:     7.776, Coord: [68,636, 53,187] degrees
Max Value: Pixel:  132.385, Coord: [82,500, -70,165] degrees
```



LSST

Stars, Milky Way, and Local Volume

Solar System

Statistics and Informatics

Dark energy

Galaxies

STRONG LENSING

Active Galactic Nuclei

Transients and Variable Stars

Large Scale Structure/Baryon Oscillation

Stars, Milky Way, and Local Volume

Solar System

Statistics and Informatics

Dark energy

Galaxies

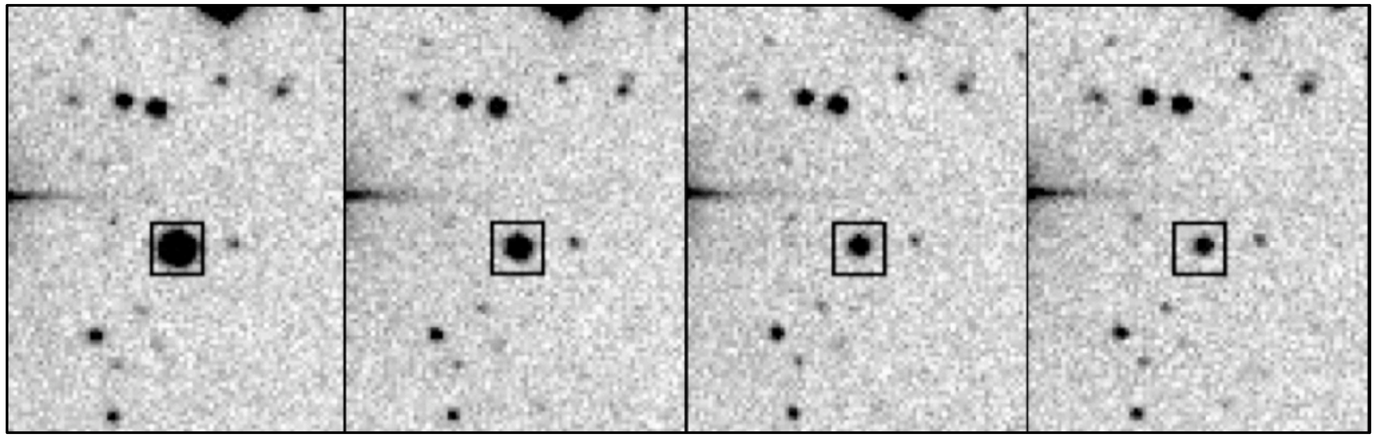
Strong Lensing

Active Galactic Nuclei

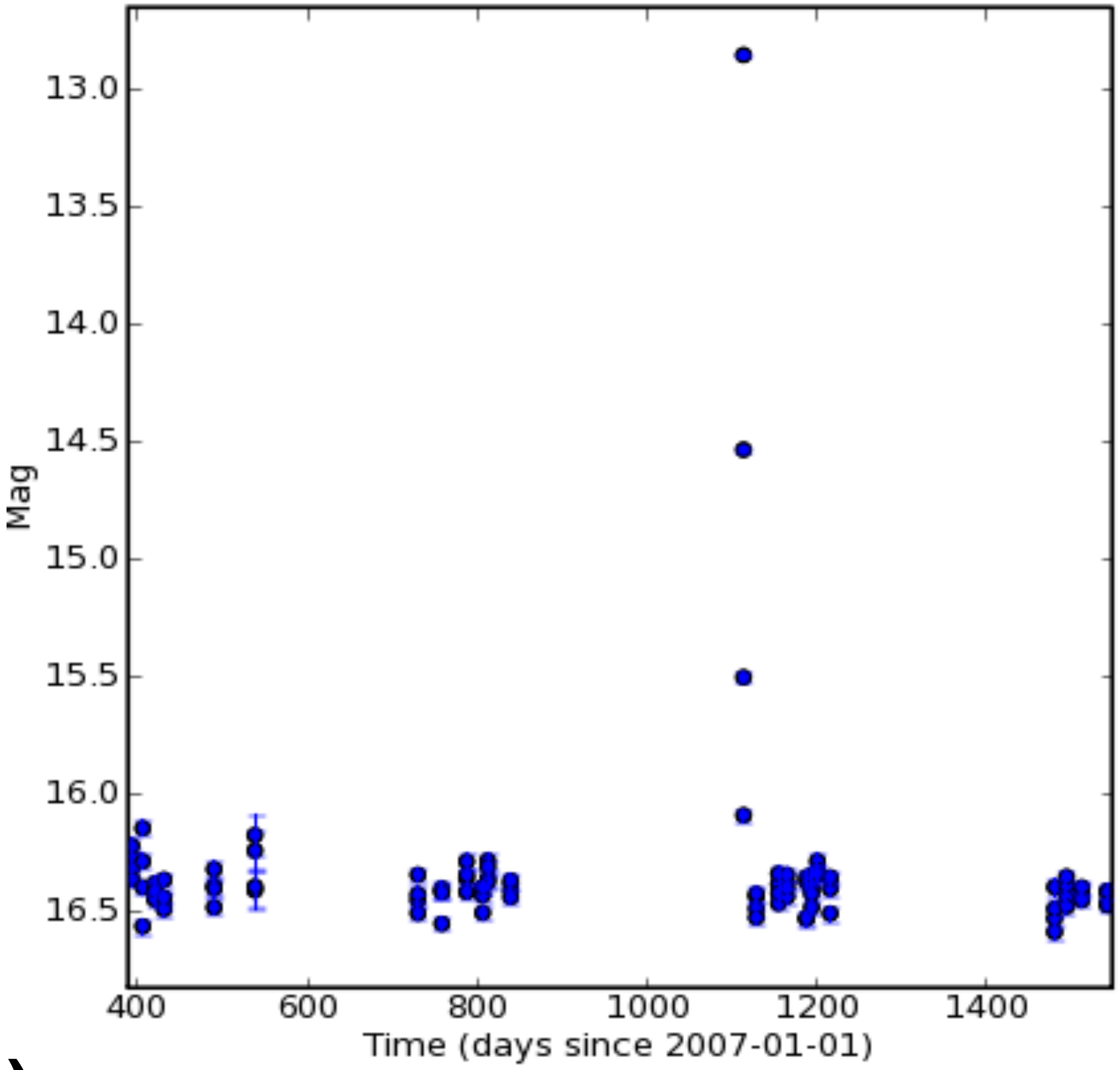
Transients and Variable Stars

Large Scale Structure/Baryon Oscillation

What is a transient?



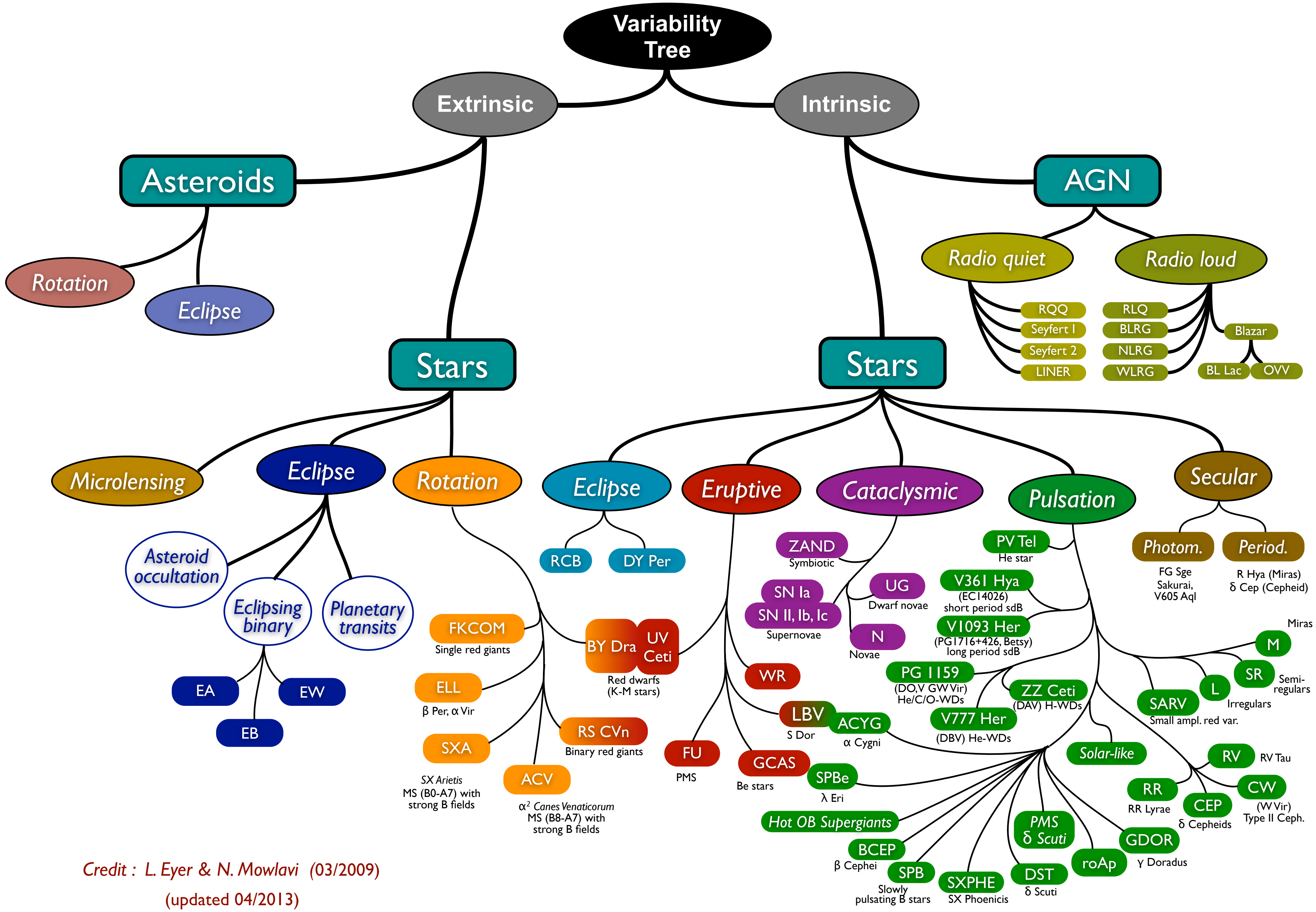
Fast transient (flaring dM), CSS080118:112149–131310



light-curve

One that has a large brightness change (delta-magnitude) within a short timespan (small delta-time)

Challenge 1: Characterize/Classify as much with as little data as possible

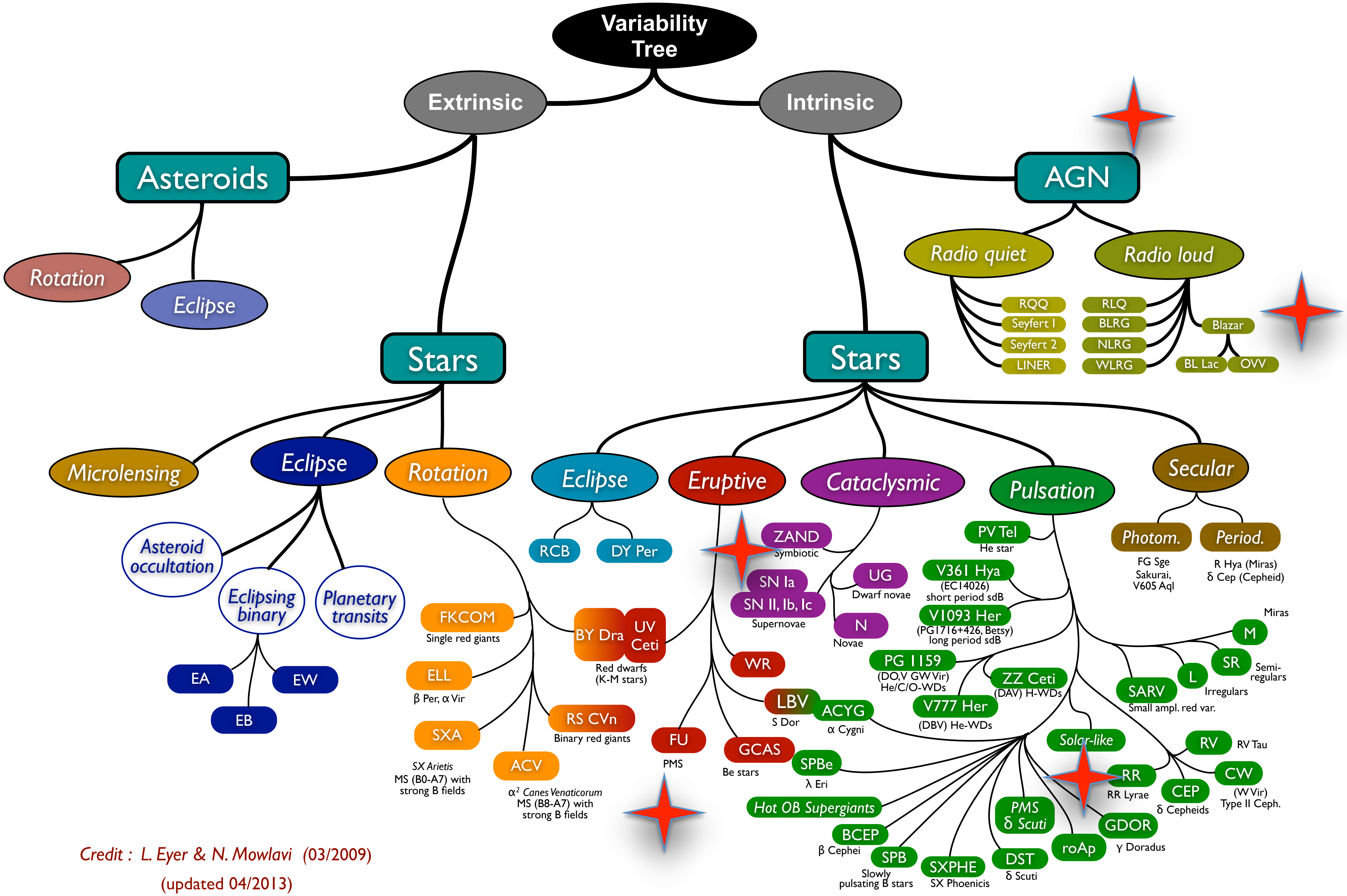


Credit : L. Eyer & N. Mowlavi (03/2009)

(updated 04/2013)

Despite the heterogeneity, gaps, heteroskedasticity

Challenge 1: Characterize/Classify as much with as little data as possible

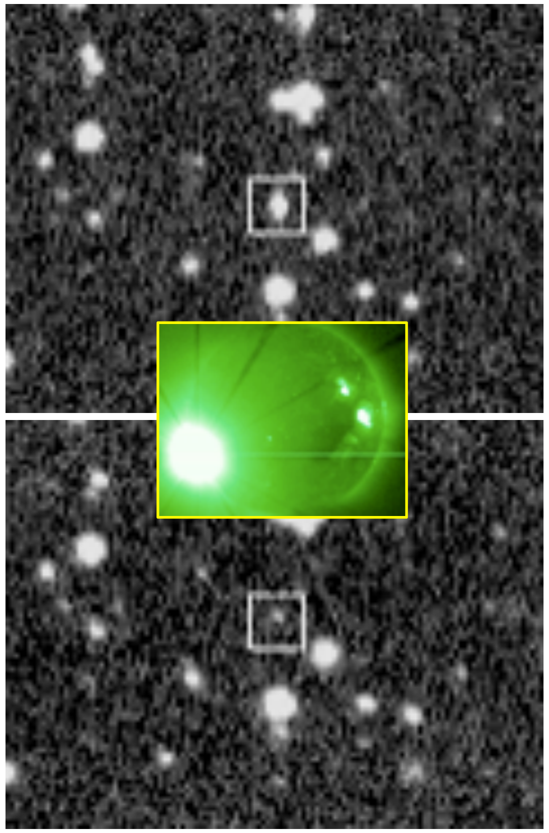


Credit : L. Eyer & N. Mowlavi (03/2009)
(updated 04/2013)

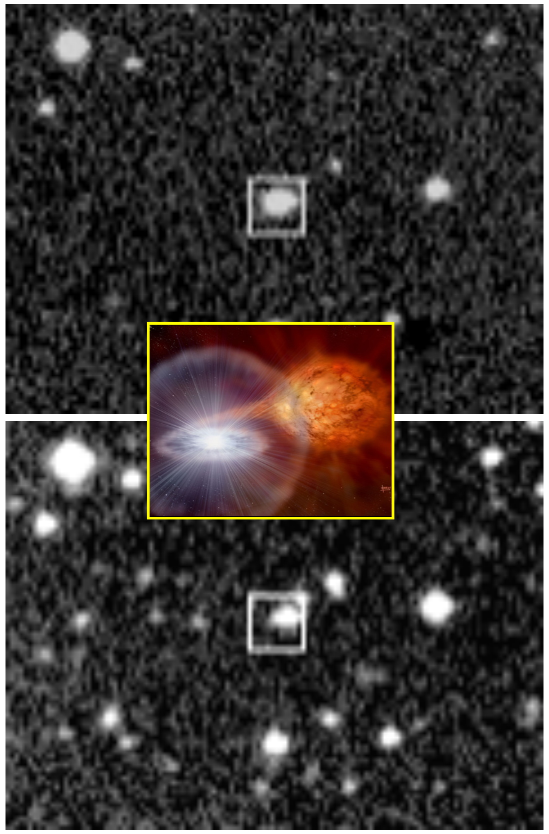
No transient left behind

Example CRTS Transients

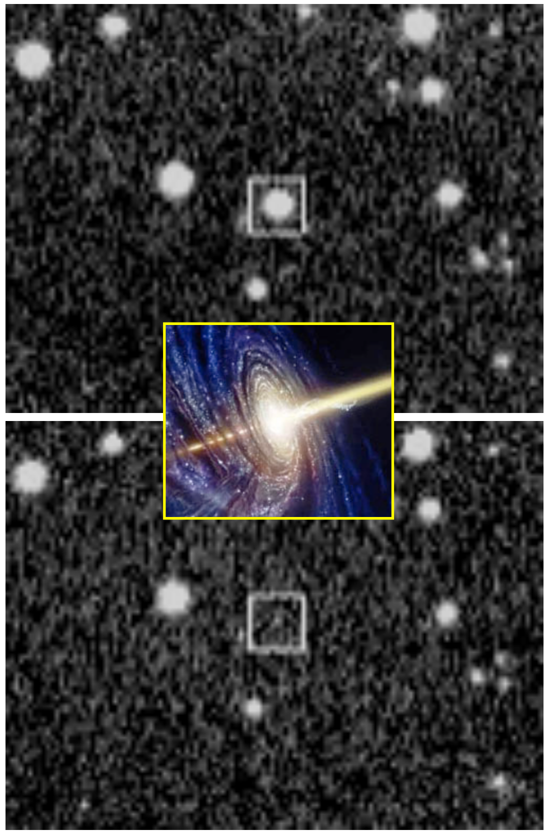
CSS090429:135125-075714
Flare star



CSS090429:101546+033311
Dwarf Nova

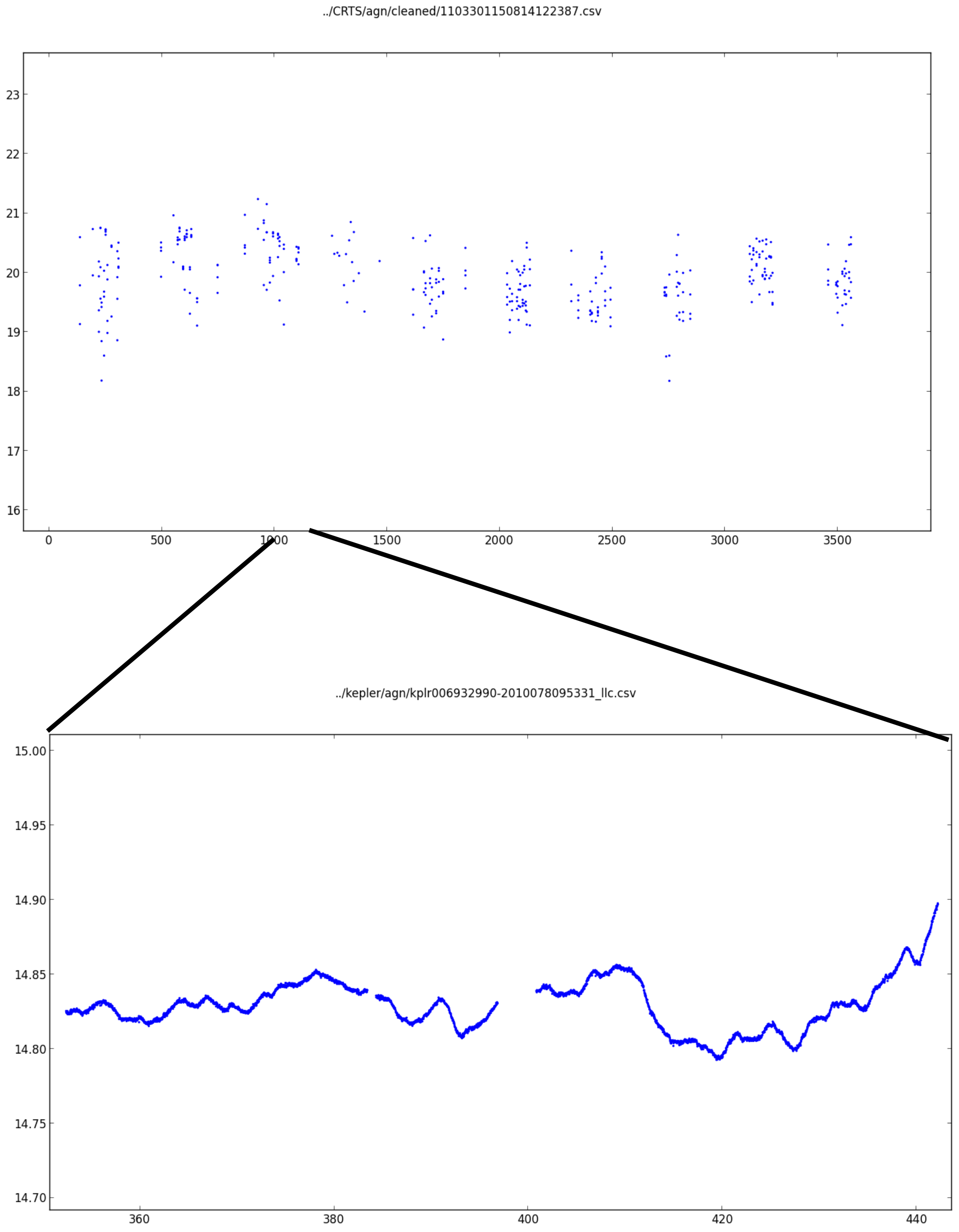


CSS090426:074240+544425
Blazar, 2EG J0744+5438



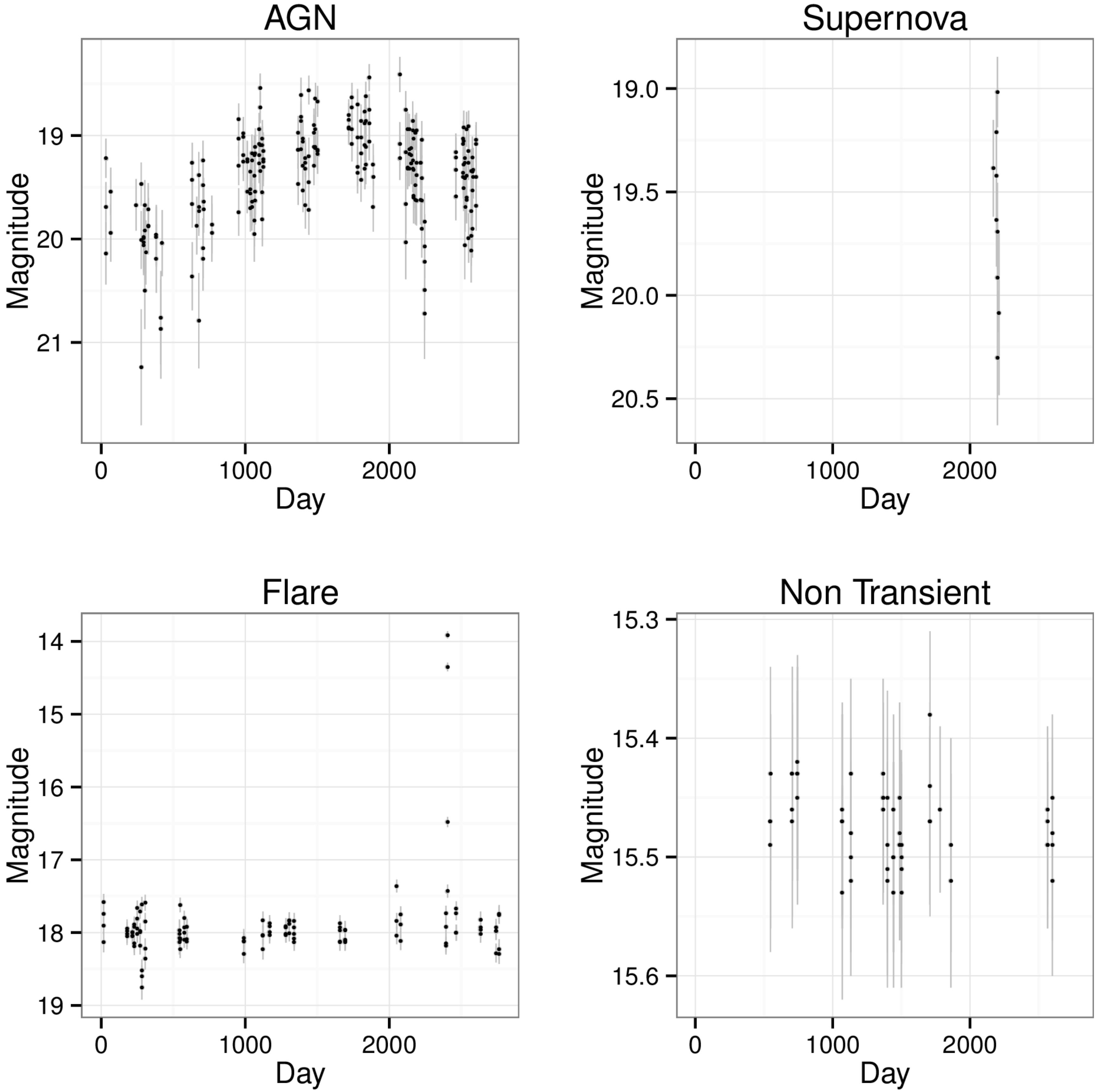
Different phenomena look the same!

AGN Variability - different perspectives

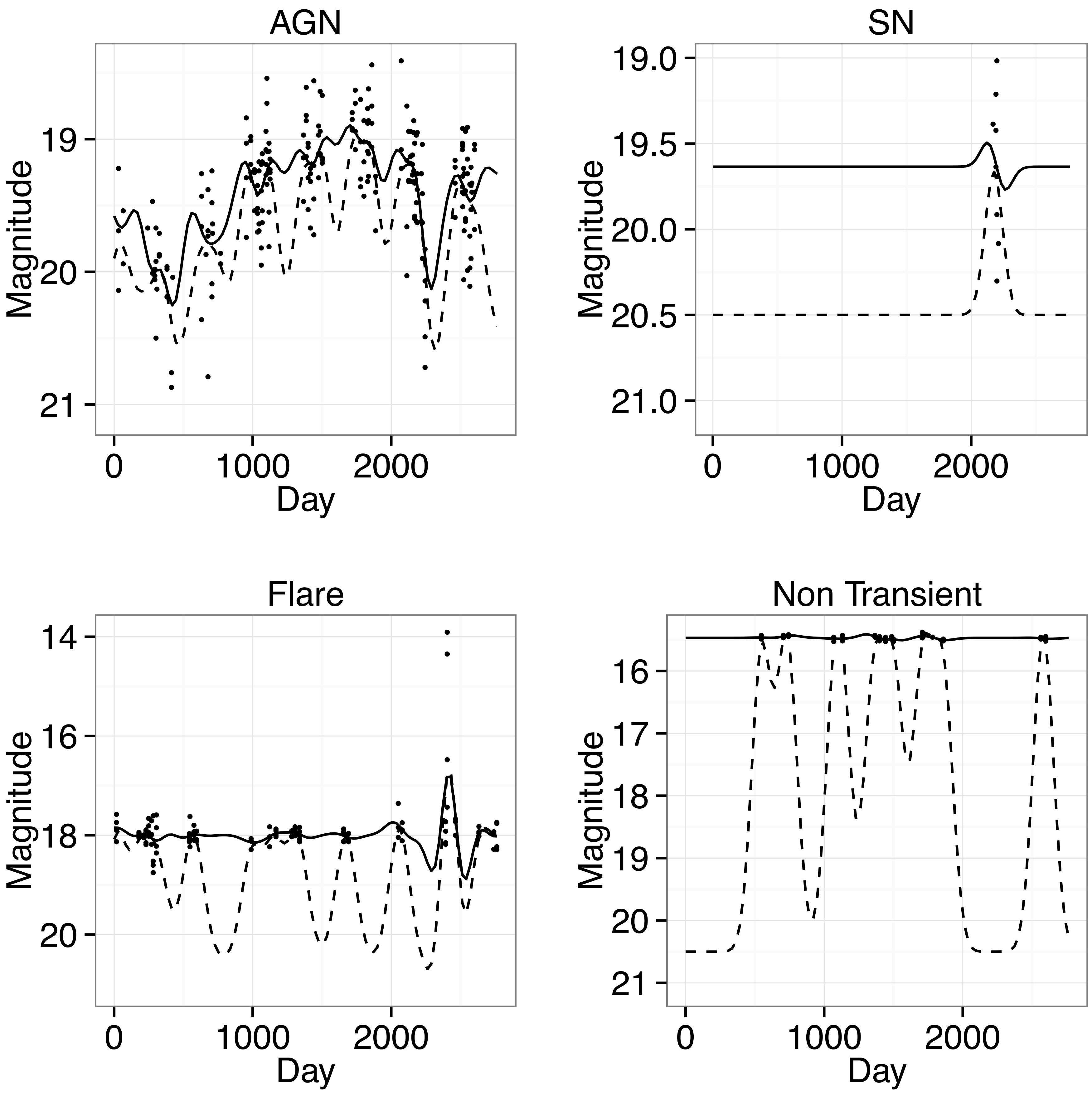


CRTS

Kepler



Truncation and Censoring

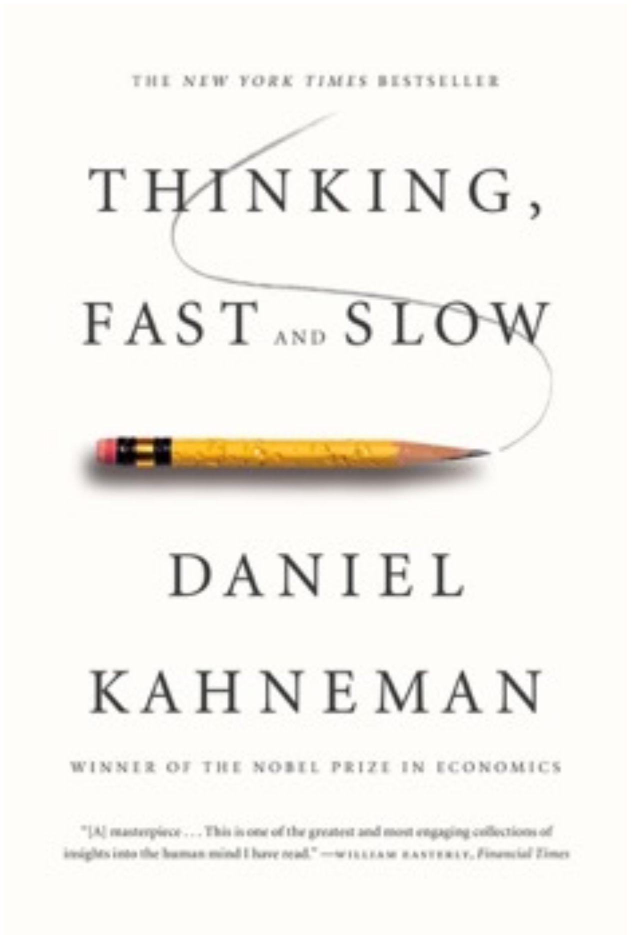


Faraway, Mahabal et al. 2015

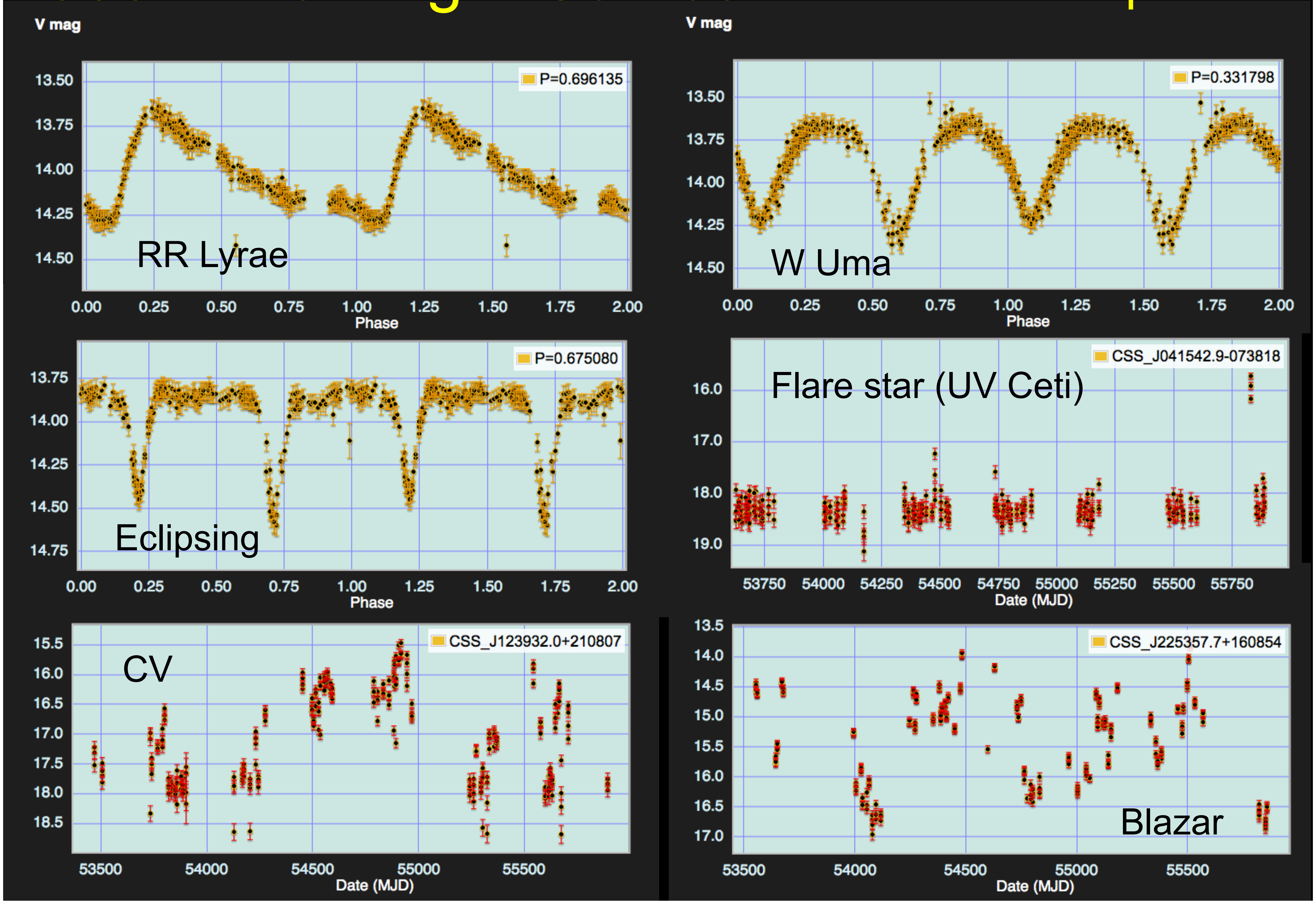
What You See Is All There Is (WYSIATI)

When regressing base rates should not be forgotten.

-


THE NEW YORK TIMES BESTSELLER
THINKING,
FAST AND SLOW
DANIEL
KAHNEMAN
WINNER OF THE NOBEL PRIZE IN ECONOMICS

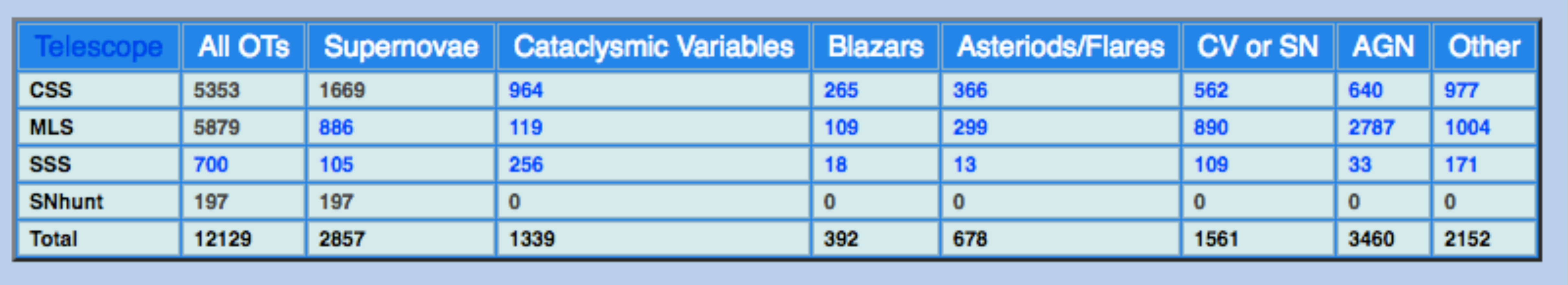
500 Million Light Curves with ~ 10^{11} data points



CRTS PIs Djorgovski, Drake

Challenge 2: Only a small fraction are rare - find/characterize them early

CRTS 10+ year status



Telescope	All OTs	Supernovae	Cataclysmic Variables	Blazars	Asteriods/Flares	CV or SN	AGN	Other
CSS	5353	1669	964	265	366	562	640	977
MLS	5879	886	119	109	299	890	2787	1004
SSS	700	105	256	18	13	109	33	171
SNhunt	197	197	0	0	0	0	0	0
Total	12129	2857	1339	392	678	1561	3460	2152

Current Status: Few tens of transients per night

Future (LSST): 10^6 - 10^7 per night; 10^4 per minute

That is why we need automatic classification algorithms

Variability on huge range of timescales

Class	Timescale	Amplitude (Δmags)
WD Pulsations	4-10 min	0.01 - 0.1
AM CVn (orbital period)	10-65 min	0.1 - 1
WD spin (int. polars)	20-60 min	0.02 - 0.4
AM CVn outbursts	1-5 days	2 - 5
Dwarf Novae outburst	4 days - 30 years	2 - 8
Symbiotic (outburst)	weeks-months	1 - 3
Novae-like high/low	days-years	2 - 5
Recurrent Novae	10-20 year	6 - 11
Novae	10^3-10^4 yr	7 - 15

Slide from Lucianne Walkowicz

Expected Rate of Transients

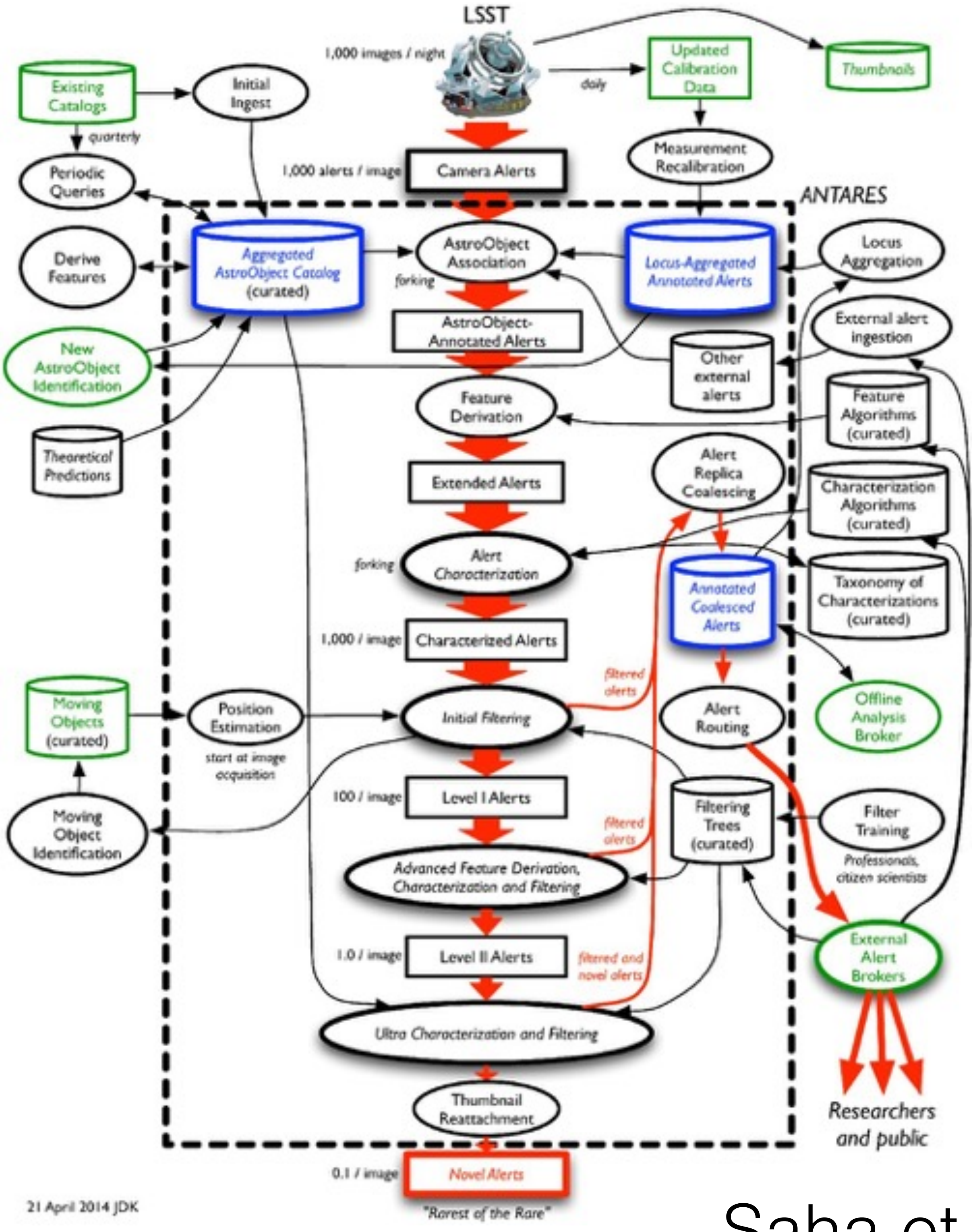
Class	Mag	t (days)	Universal Rate	**LSST Rate**
Luminous SNe	-19...-23	50 - 400	10^{-7} Mpc^{-3} yr^{-1}	**20000**
Orphan Afterglows SHB	-14...-18	5 -15	3 x$10^{-7...-9}$ Mpc^{-3} yr^{-1}	**~10 - 100**
Orphan Afterglows LSB	-22...-26	2 - 15	3 x $10^{-10...-11}$ Mpc^{-3} yr^{-1}	**1000**
On-axis GRB afterglows	...-37	1 - 15	10^{-11} Mpc^{-3} yr^{-1}	**~50**
Tidal Disruption Flares	-15...-19	30 - 350	10^{-6} Mpc^{-3} yr^{-1}	**6000**
Luminous Red Novae	-9...-13	20 - 60	10^{-13} yr^{-1} Lsun^{-1}	**80 - 3400**
Fallback SNe	-4...-21	0.5 - 2	<5 x 10^{-6} Mpc^{-3} yr^{-1}	**< 800**
SNe Ia	-17...-19.5	30 - 70	3 x 10^{-5} Mpc^{-3} yr^{-1}	**200000**
SNe II	-15...-20	20 - 300	(3..8) x 10^{-5} Mpc^{-3} yr^{-1}	**100000**

Table adapted from Rau et al. 2009 by Lucianne Walkowicz

NOAO’s proposed broker Antares

- Solar System
- LSST History
- Other catalogs
- Ancillary data

0.1 **rare** alerts/image



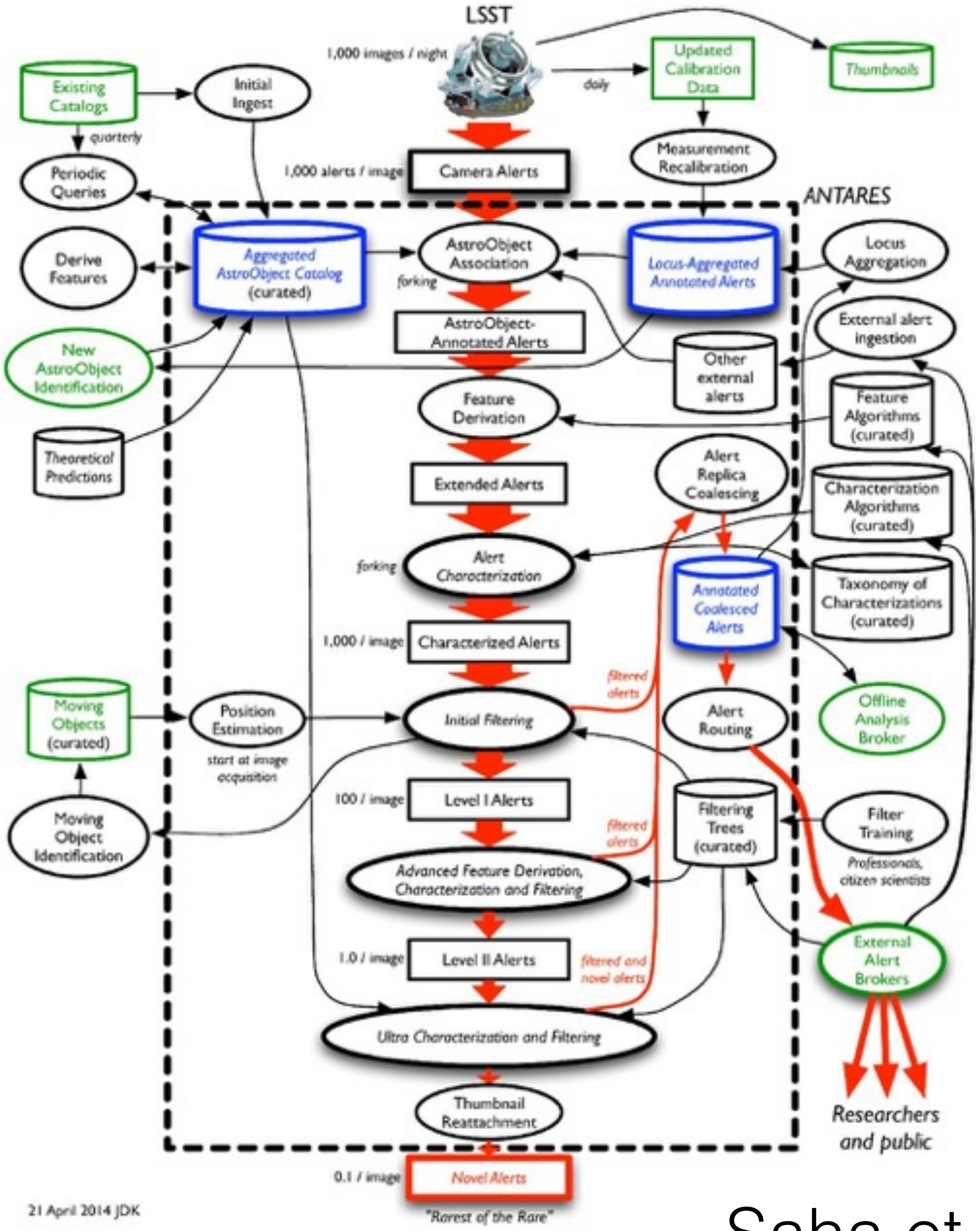
Saha et al
1409.0056

NOAO's proposed broker Antares

- Solar System
- LSST History
- Other catalogs
- Ancillary data

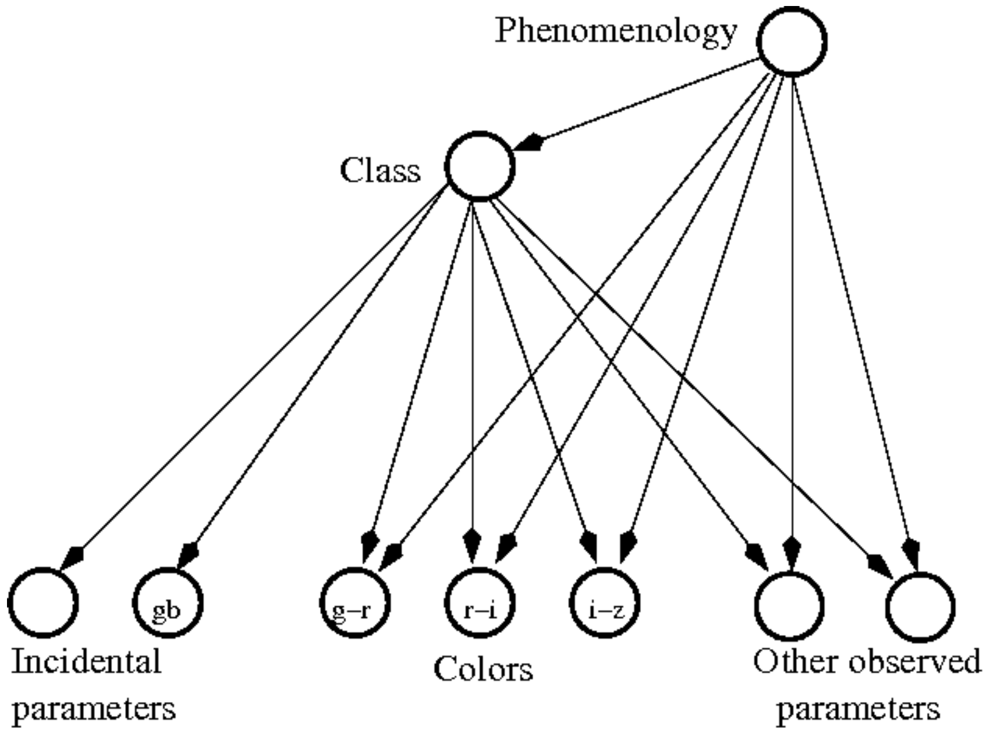
0.1 rare alerts/image

2017 workshop
2016 LSST AHM

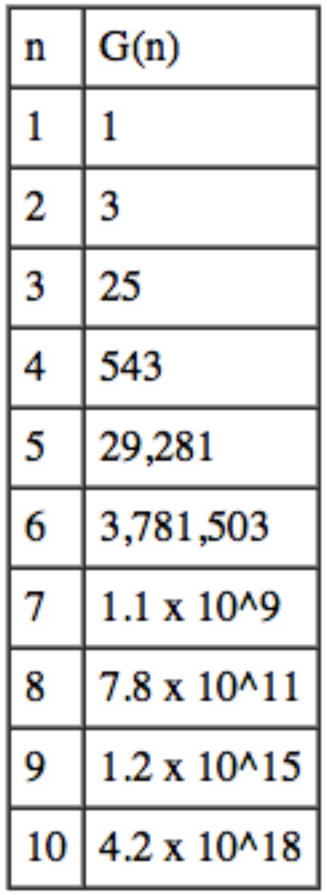


Saha et al
1409.0056

Bayesian Networks



Search space growth hyperexponential

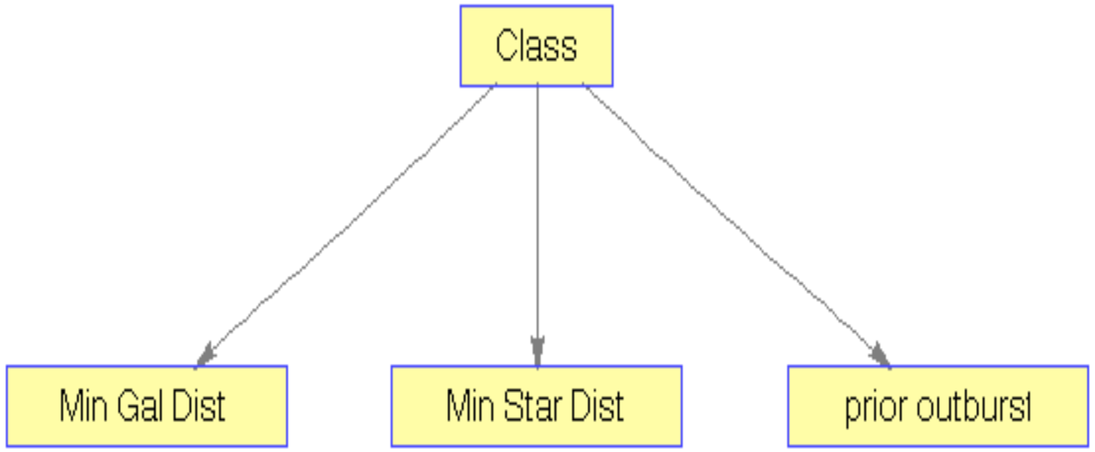


n	G(n)
1	1
2	3
3	25
4	543
5	29,281
6	3,781,503
7	1.1 x 10^9
8	7.8 x 10^11
9	1.2 x 10^15
10	4.2 x 10^18

Very broadly speaking 5 flavors of BNs possible

- Naïve
- Tree Augmented Network (TAN)
- **Constructed (semantics, expert knowledge etc. based)**
- **Single winner from several naïve**
- **Fully learned from data**

SNe/non-SNe BN

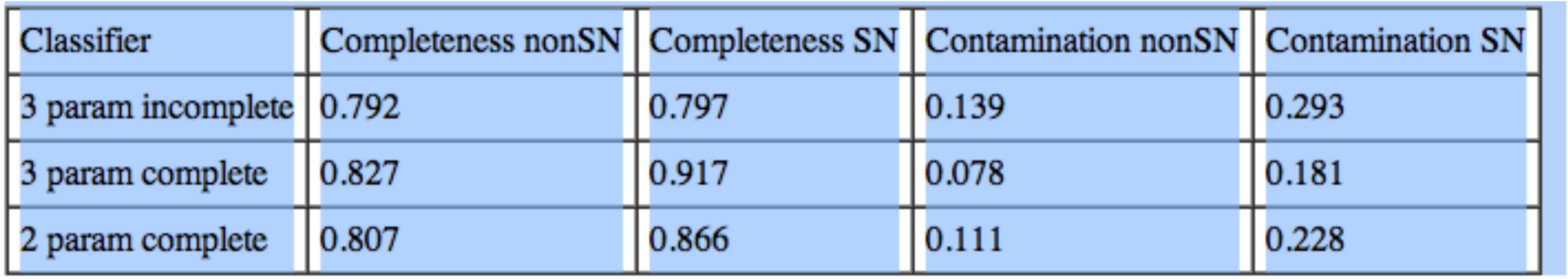


normalized

Based on peaks

$$\text{prior outburst} = \frac{1}{t_{span}} \cdot \left(\frac{\sum_i w_i (p_i - p_m)^2}{N}\right)^{1/2}$$

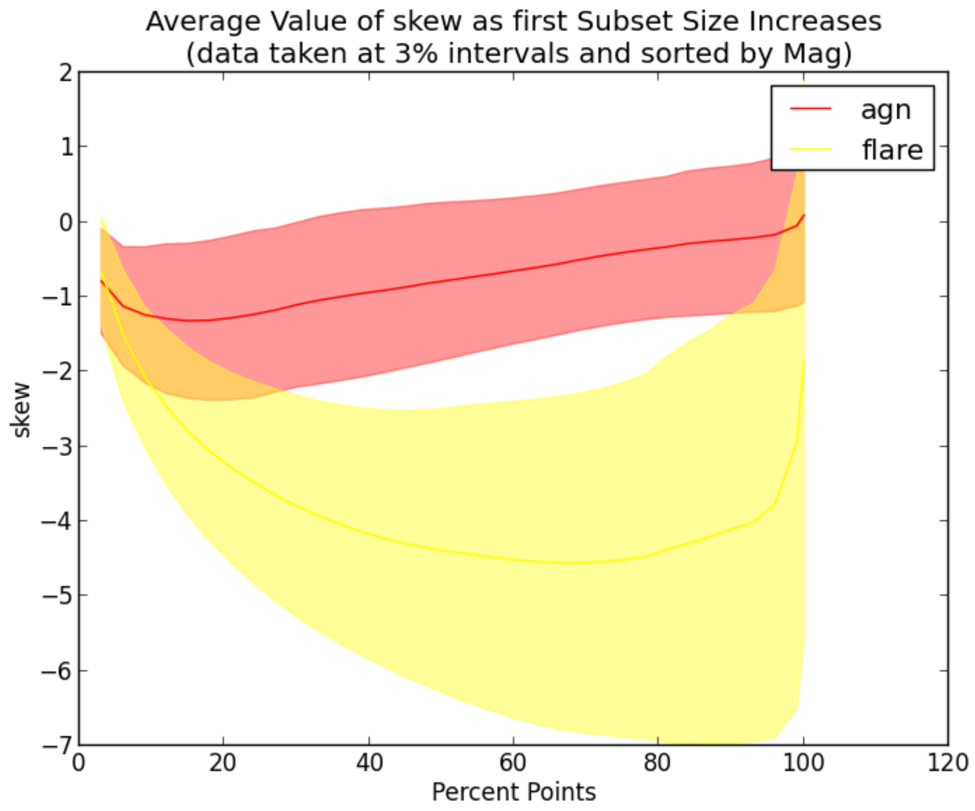
80-90% completeness
Only archival information

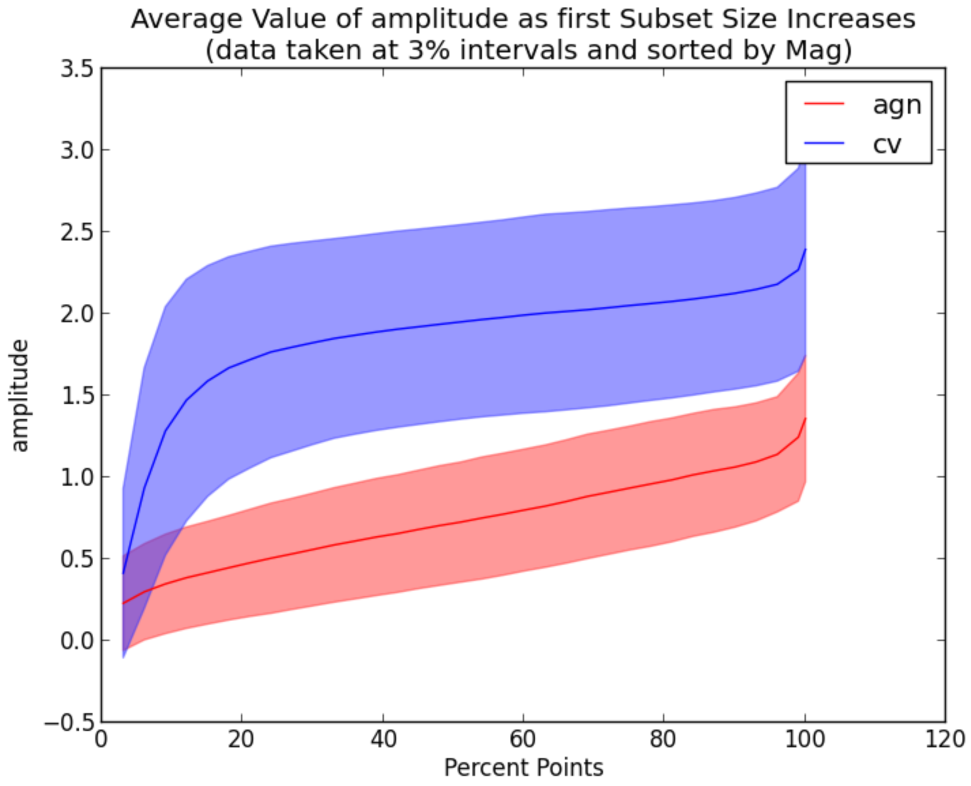


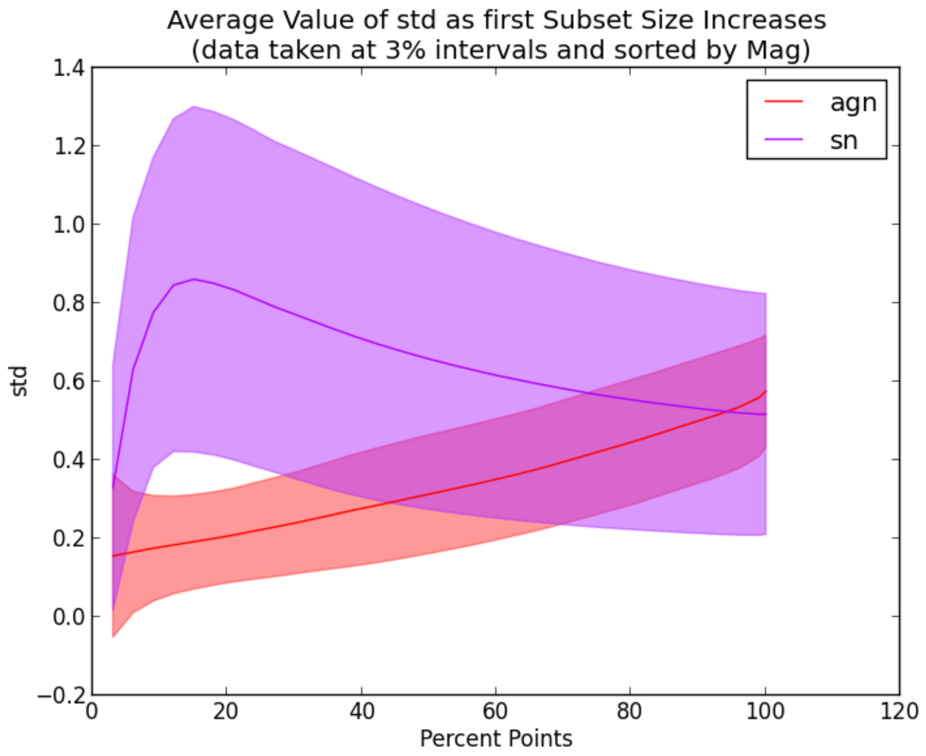
Classifier	Completeness nonSN	Completeness SN	Contamination nonSN	Contamination SN
3 param incomplete	0.792	0.797	0.139	0.293
3 param complete	0.827	0.917	0.078	0.181
2 param complete	0.807	0.866	0.111	0.228

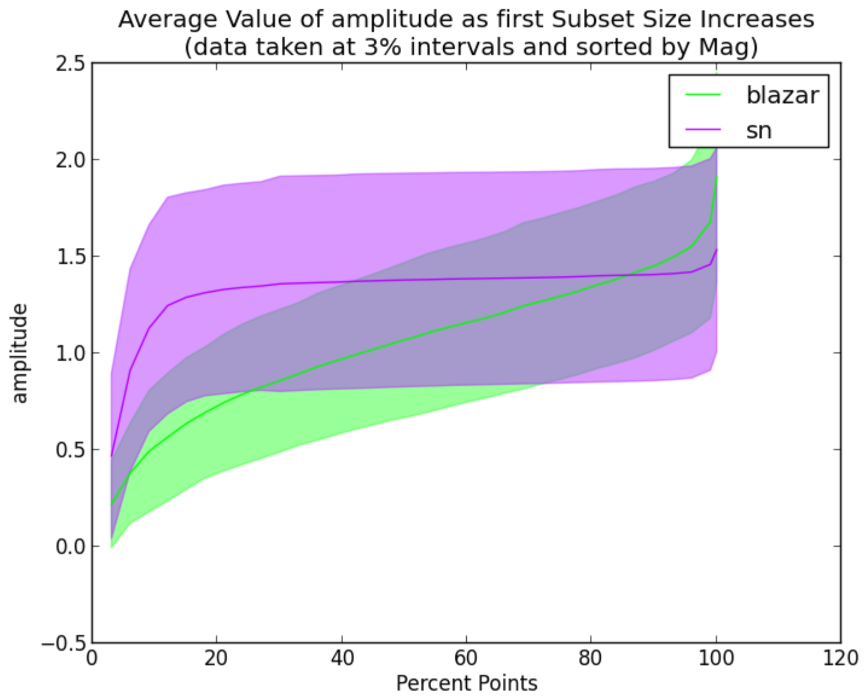
Discriminating features

Chengyi Lee


Average Value of skew as first Subset Size Increases
(data taken at 3% intervals and sorted by Mag)

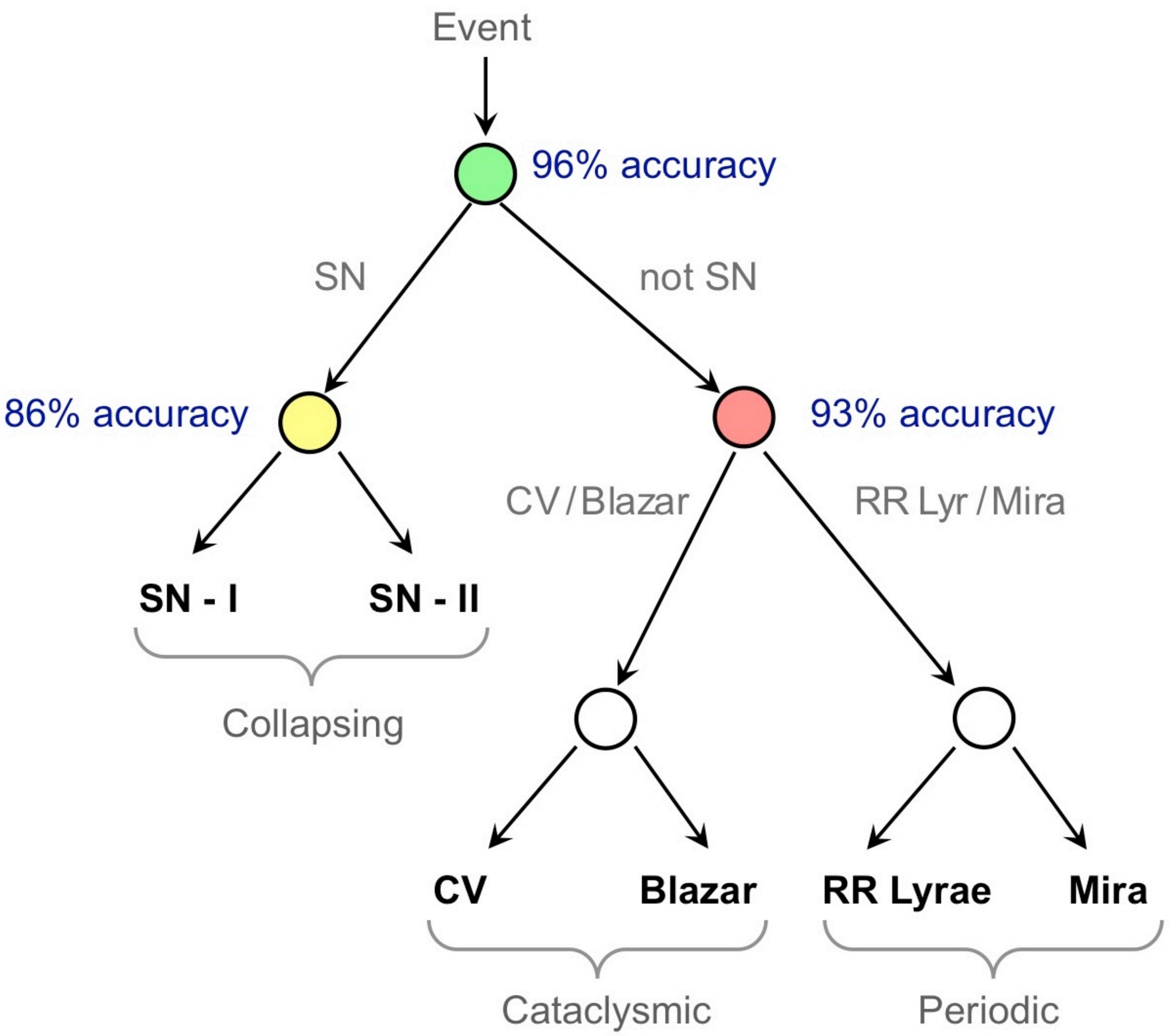

Average Value of amplitude as first Subset Size Increases
(data taken at 3% intervals and sorted by Mag)


Average Value of std as first Subset Size Increases
(data taken at 3% intervals and sorted by Mag)

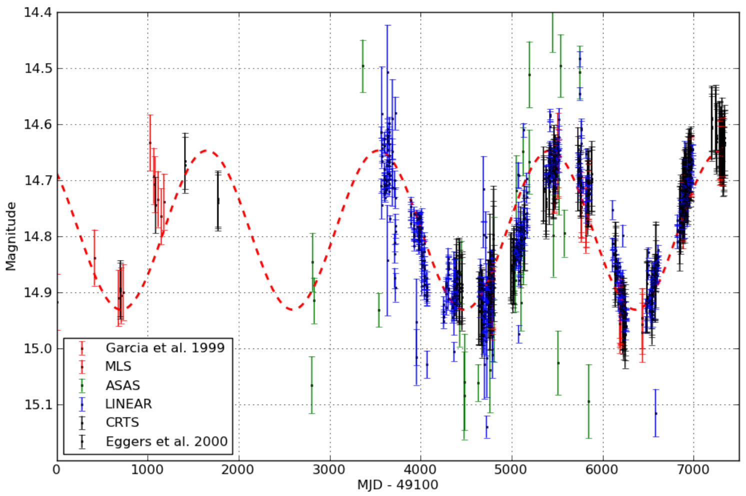

Average Value of amplitude as first Subset Size Increases
(data taken at 3% intervals and sorted by Mag)

You can not step into the same river twice.

Hierarchical approach



Archival search

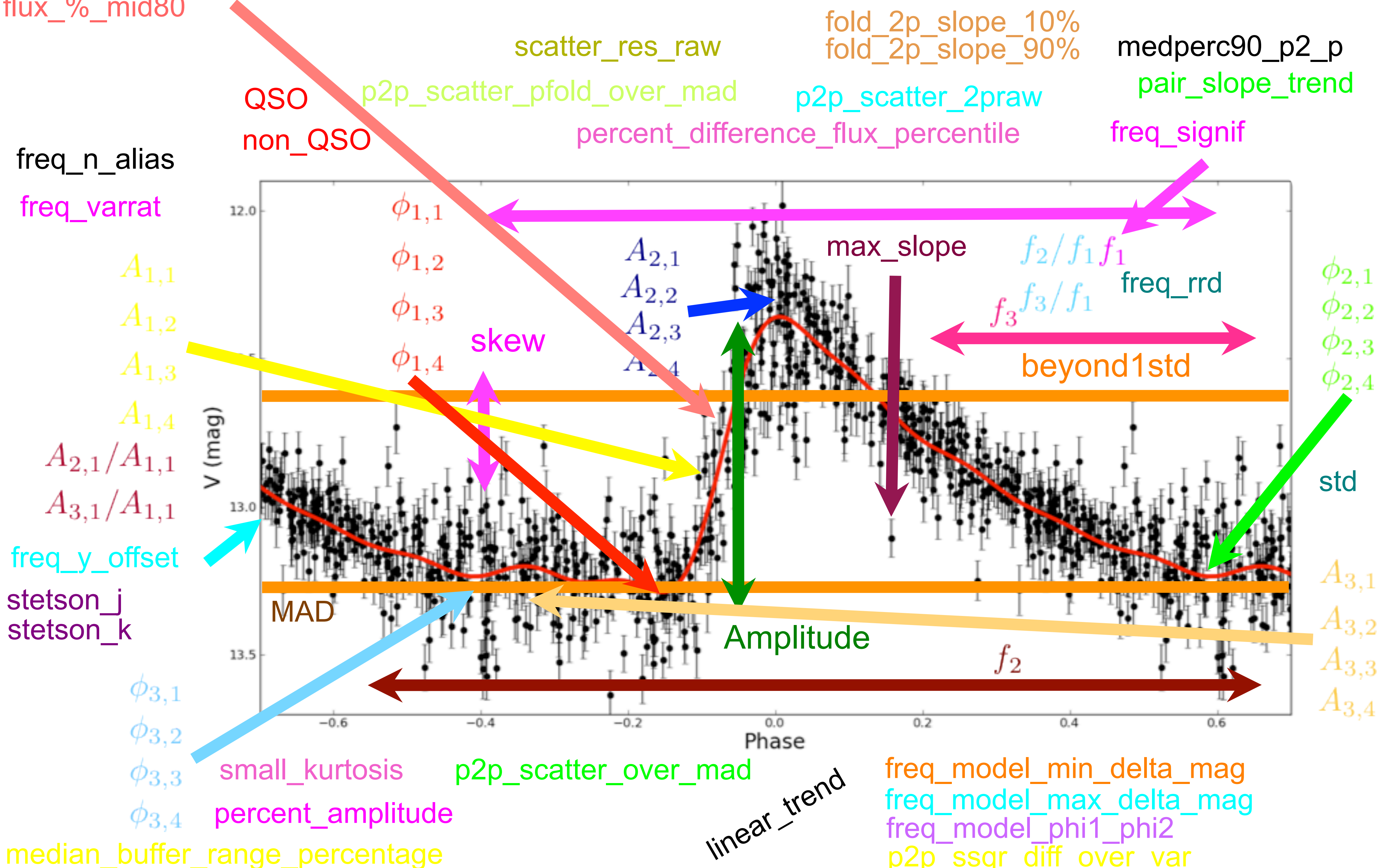


Binary Blackholes
Graham et al. 2015
CARMA/Wavelets

Many features
- not all are independent

Adam Miller

flux_%_mid20
flux_%_mid35
flux_%_mid50
flux_%_mid65
flux_%_mid80



scatter_res_raw

fold_2p_slope_10%
fold_2p_slope_90%

medperc90_p2_p

pair_slope_trend

QSO

non_QSO

p2p_scatter_pfold_over_mad

p2p_scatter_2praw

percent_difference_flux_percentile

freq_signif

freq_n_alias

freq_varrat

$\phi_{1,1}$ $\phi_{1,2}$ $\phi_{1,3}$ $\phi_{1,4}$

$A_{1,1}$ $A_{1,2}$ $A_{1,3}$ $A_{1,4}$

$A_{2,1}$ $A_{2,2}$ $A_{2,3}$ $A_{2,4}$

max_slope

f_2/f_1 f_1

freq_rrd

f_3 f_3/f_1

$\phi_{2,1}$ $\phi_{2,2}$ $\phi_{2,3}$ $\phi_{2,4}$

skew

beyond1std

$A_{2,1}/A_{1,1}$

$A_{3,1}/A_{1,1}$

std

freq_y_offset

stetson_j
stetson_k

MAD

Amplitude

$A_{3,1}$ $A_{3,2}$ $A_{3,3}$ $A_{3,4}$

f_2

$\phi_{3,1}$ $\phi_{3,2}$ $\phi_{3,3}$ $\phi_{3,4}$

small_kurtosis

p2p_scatter_over_mad

linear_trend

freq_model_min_delta_mag
freq_model_max_delta_mag
freq_model_phi1_phi2
p2p_ssqr_diff_over_var

percent_amplitude

median_buffer_range_percentage

Resort to dimensionality reduction

Challenge 3: A variety of parameters - choose judiciously

Discovery; Contextual; Follow-up; Prior Classification …

Whole curve measures

Median magnitude (mag); mean of absolute differences of successive observed magnitude; the maximum difference magnitudes

Fitted curve measures

Scaled total variation scaled by number of days of observation; range of fitted curve; maximum derivative in the fitted curve

Residual from fit measures

The maximum studentized residual; SD of residuals; skewness of residuals; Shapiro-Wilk statistic of residuals

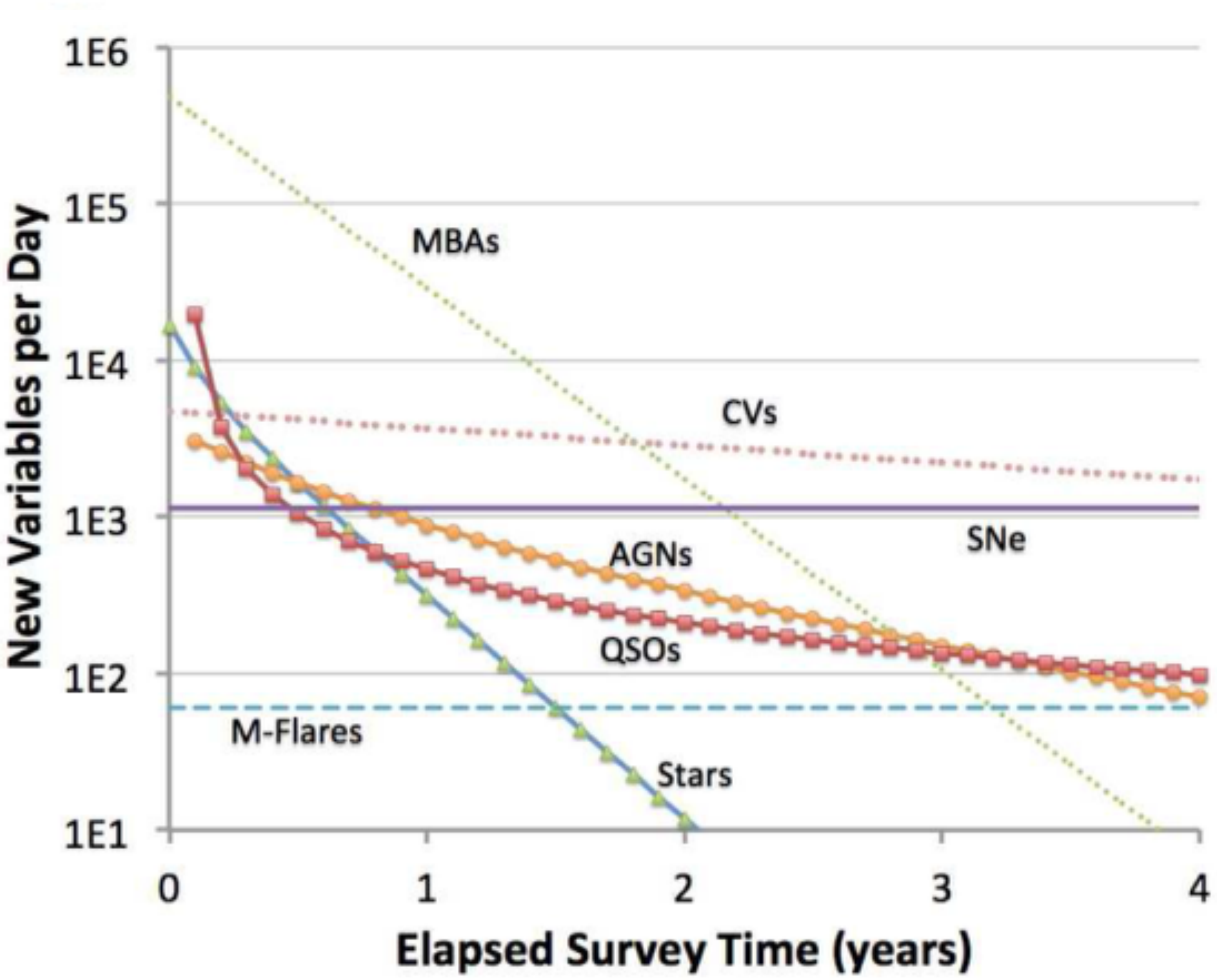
Cluster measures

Fit the means within the groups (up to 4 measurements); and then take the logged SD of the residuals from this fit; the max absolute residuals from this fit; total variation of curve based on group means scaled by range of observation

Challenge 4: real-time computation required - find ways to make that happen

Recomputation of features

Updating priors

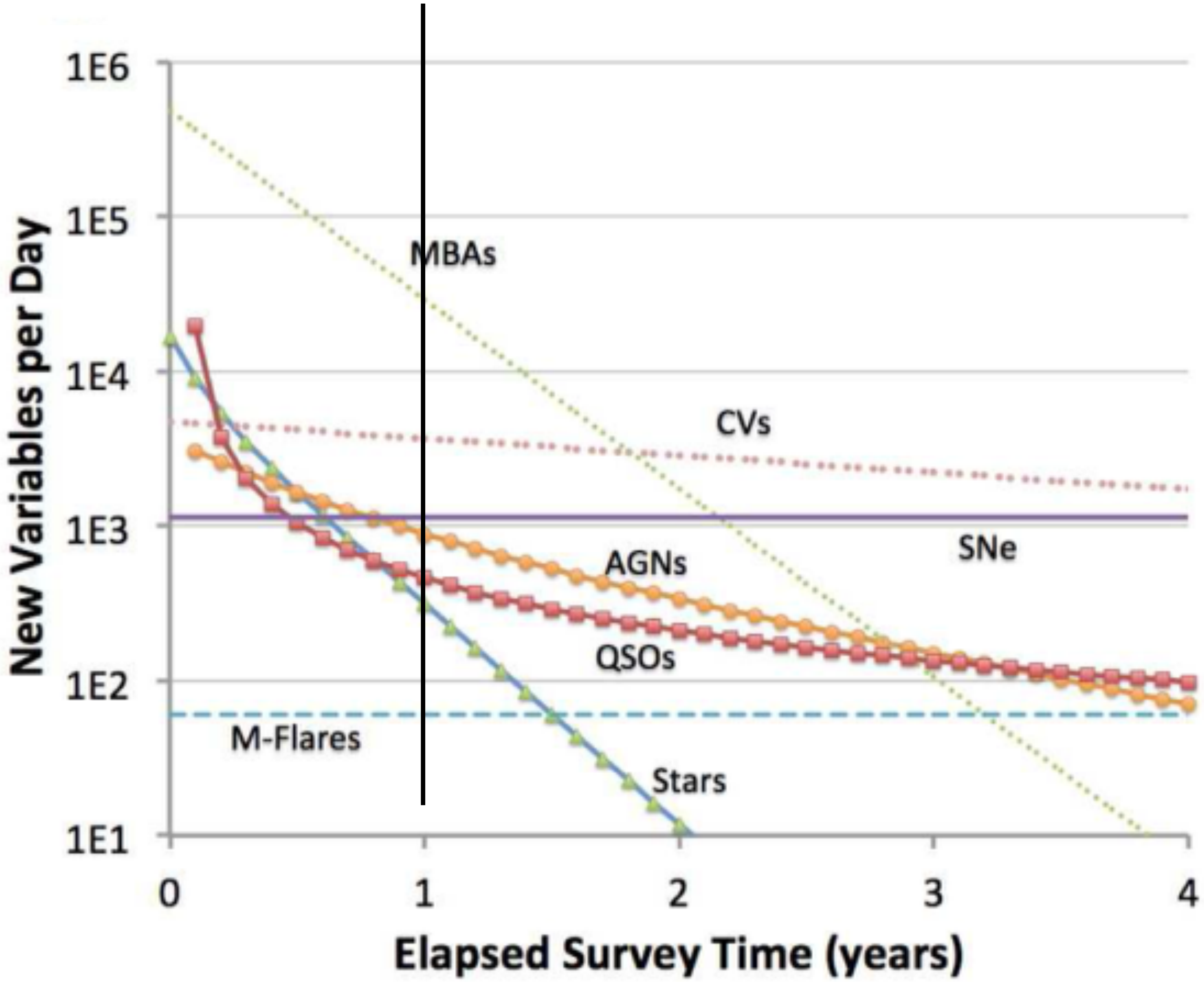


Ridgeway et al., arXiv: 1409.3265

Challenge 4: real-time computation required - find ways to make that happen

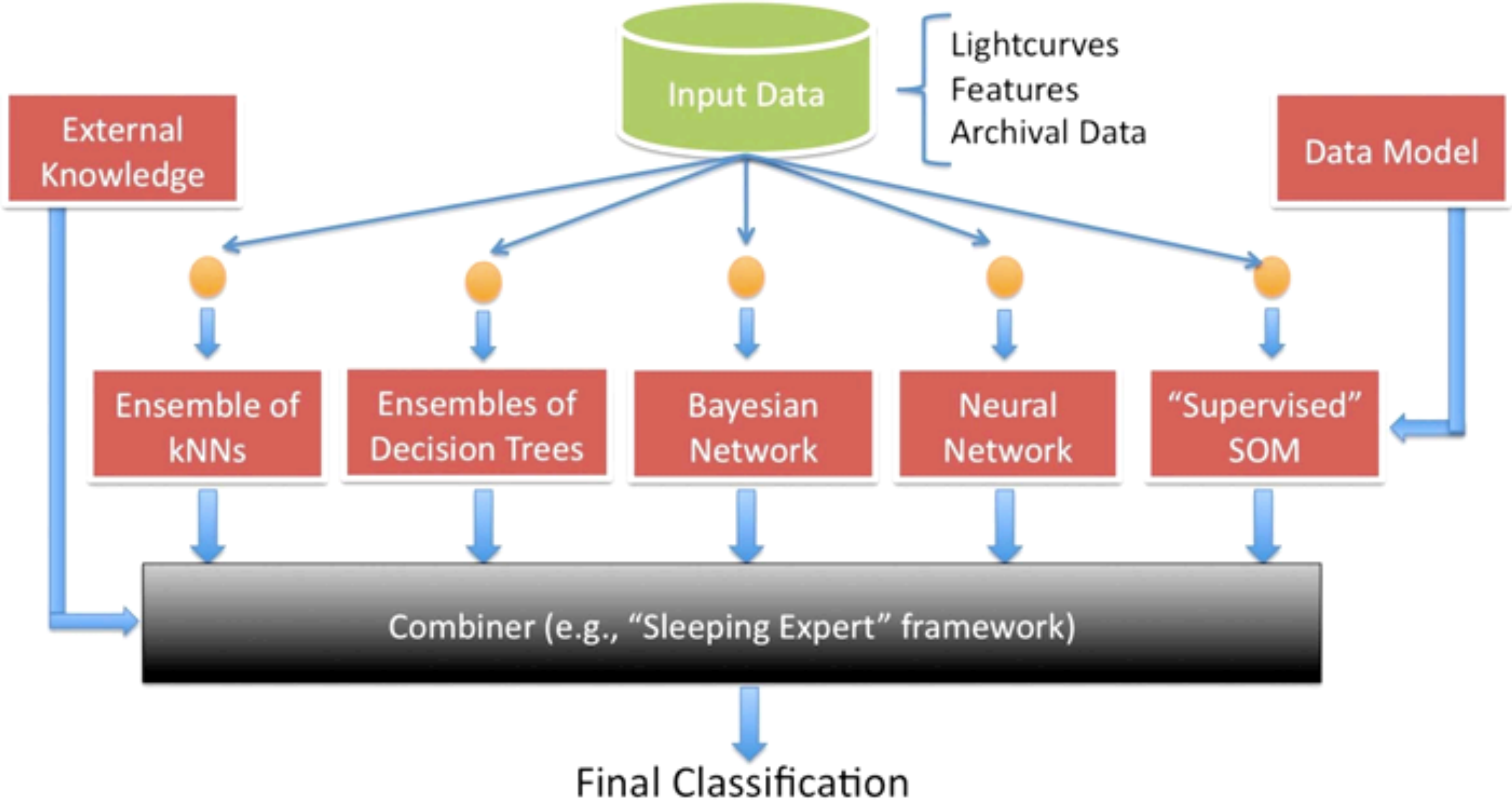
Recomputation of features

Updating priors



Ridgeway et al., arXiv: 1409.3265

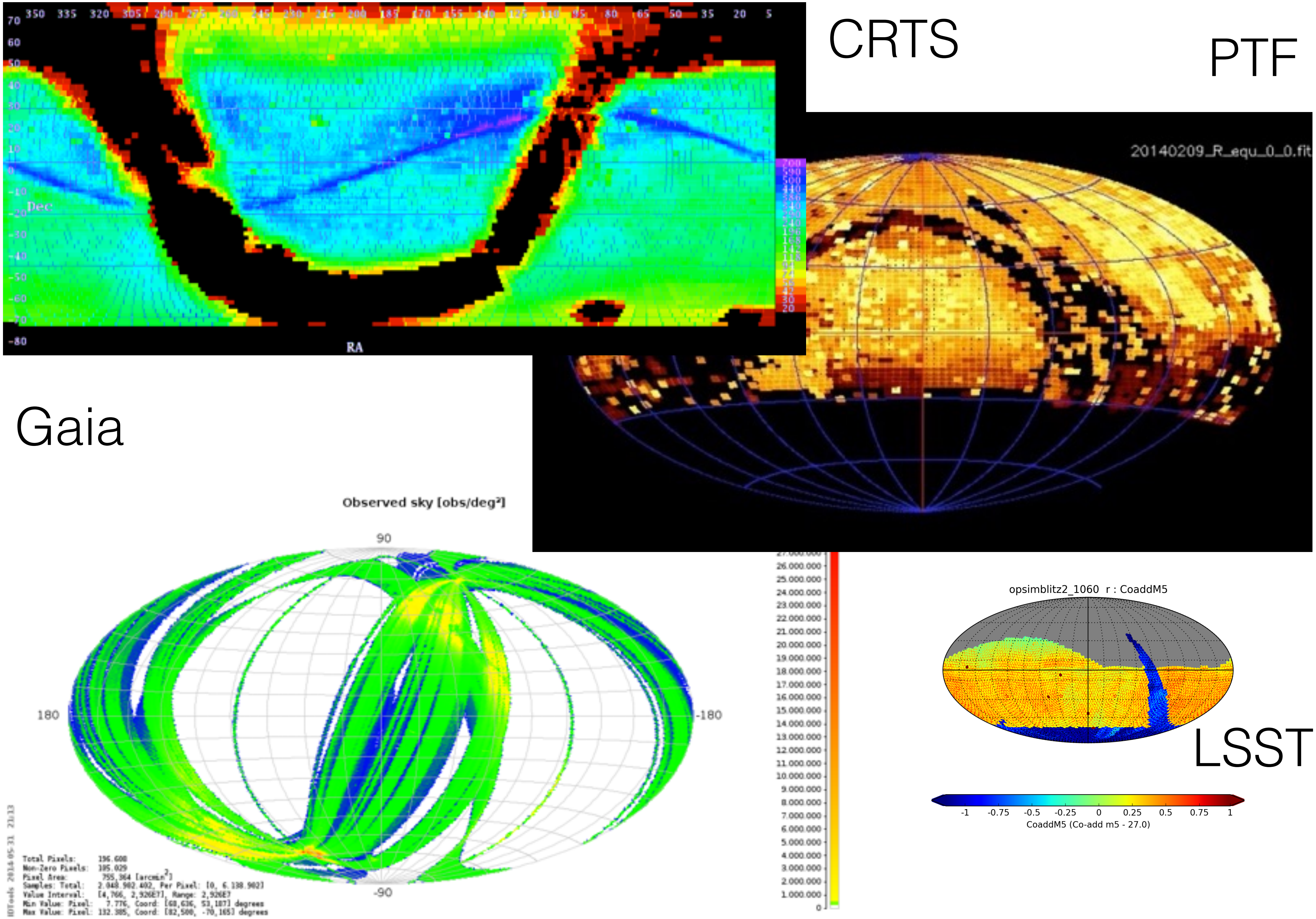
Challenge 5: Metaclassification - combining diverse classifiers optimally



As varied classifiers are used for parts of the classification tree combing their outputs in an optimal way becomes crucial

Mahabal, Donalek

Sky Maps of a few surveys



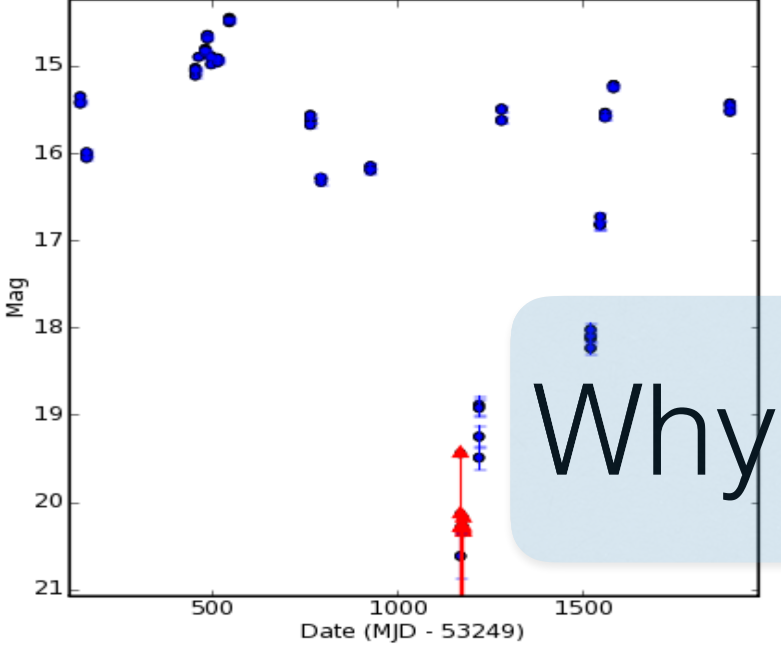
CRTS

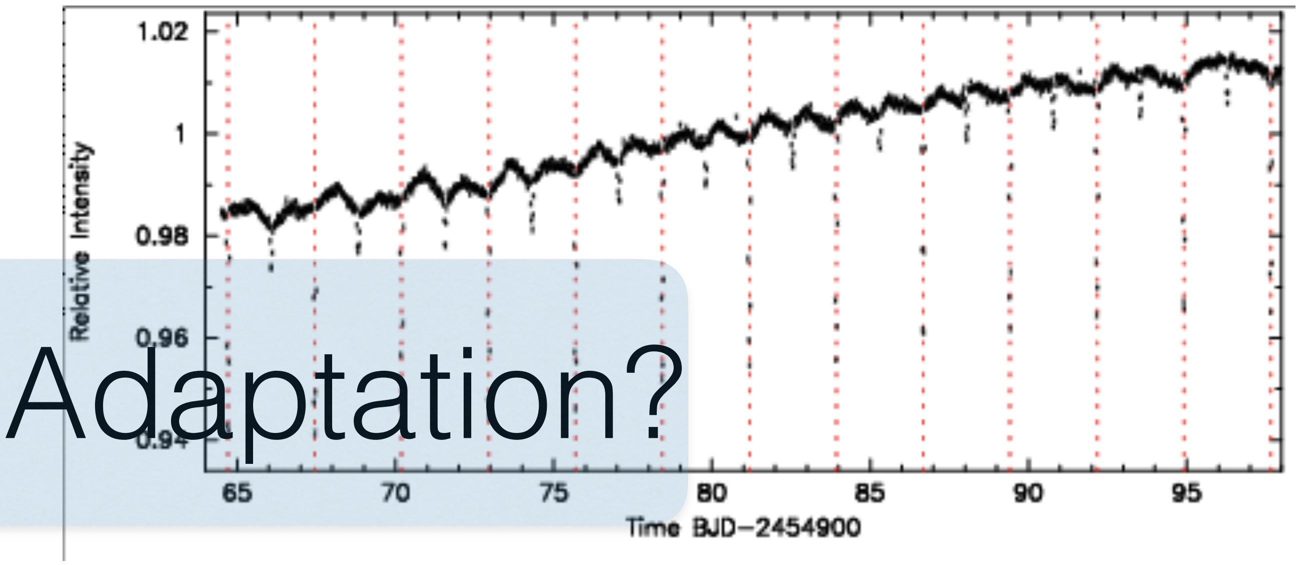
PTF

Gaia

LSST

```
Total Pixels:       196.608
Non-Zero Pixels:    105.029
Pixel Area:          755,364 [arcmin²]
Samples: Total:     2.048.902.402, Per Pixel: [0, 6.138.902]
Value Interval:     [4,766, 2,926E7], Range: 2,926E7
Min Value: Pixel:    7.776, Coord: [68,636, 53,187] degrees
Max Value: Pixel:   132.385, Coord: [82,500, -70,165] degrees
```



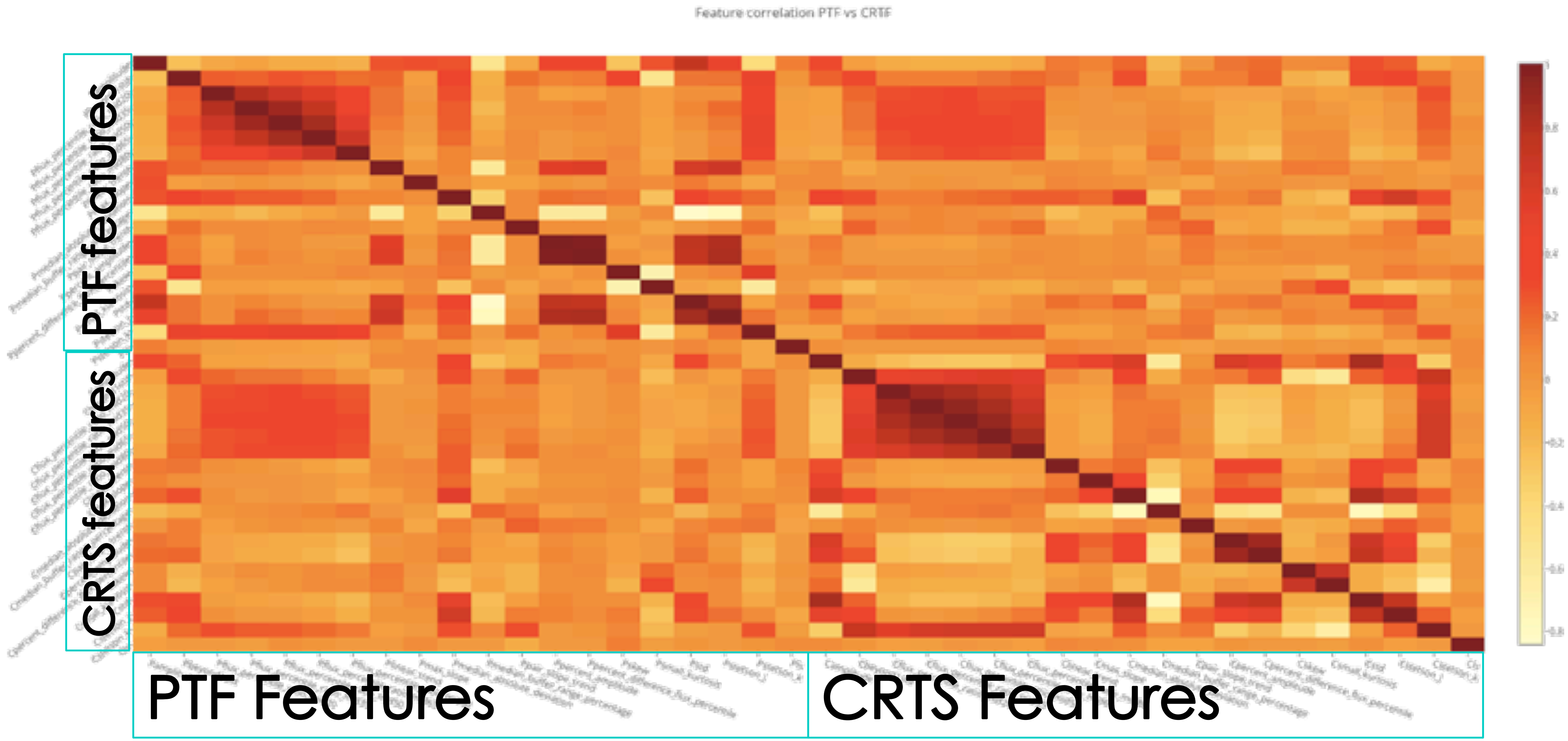


Why Domain Adaptation?

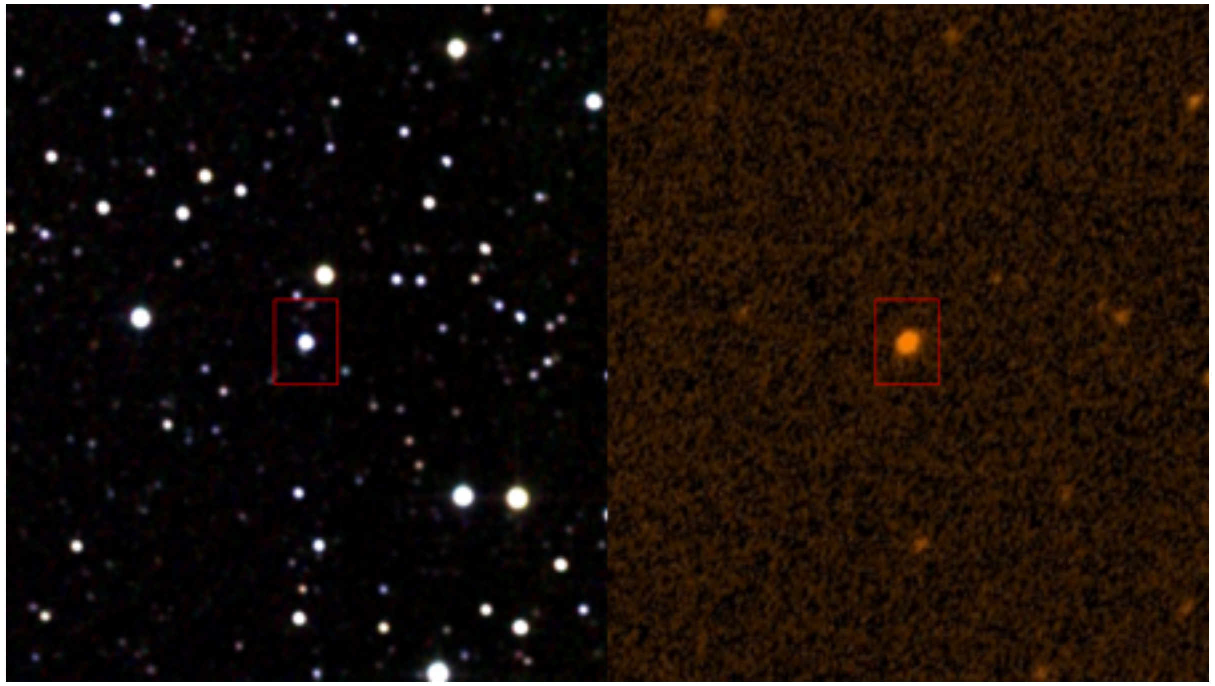
- Surveys differ in depth (aperture), filters, cadence
- Same (type of) objects produce different statistical features (skew, median absolute deviation etc.)
- Learning tends to be done on each survey separately - leading to unnecessary delays
- DA helps build on the otherwise untapped intersurvey synergy (think DASCH -> CRTS/ZTF/Kepler -> LSST)

Jingling Li, S Vaijanapurkar, B Bue

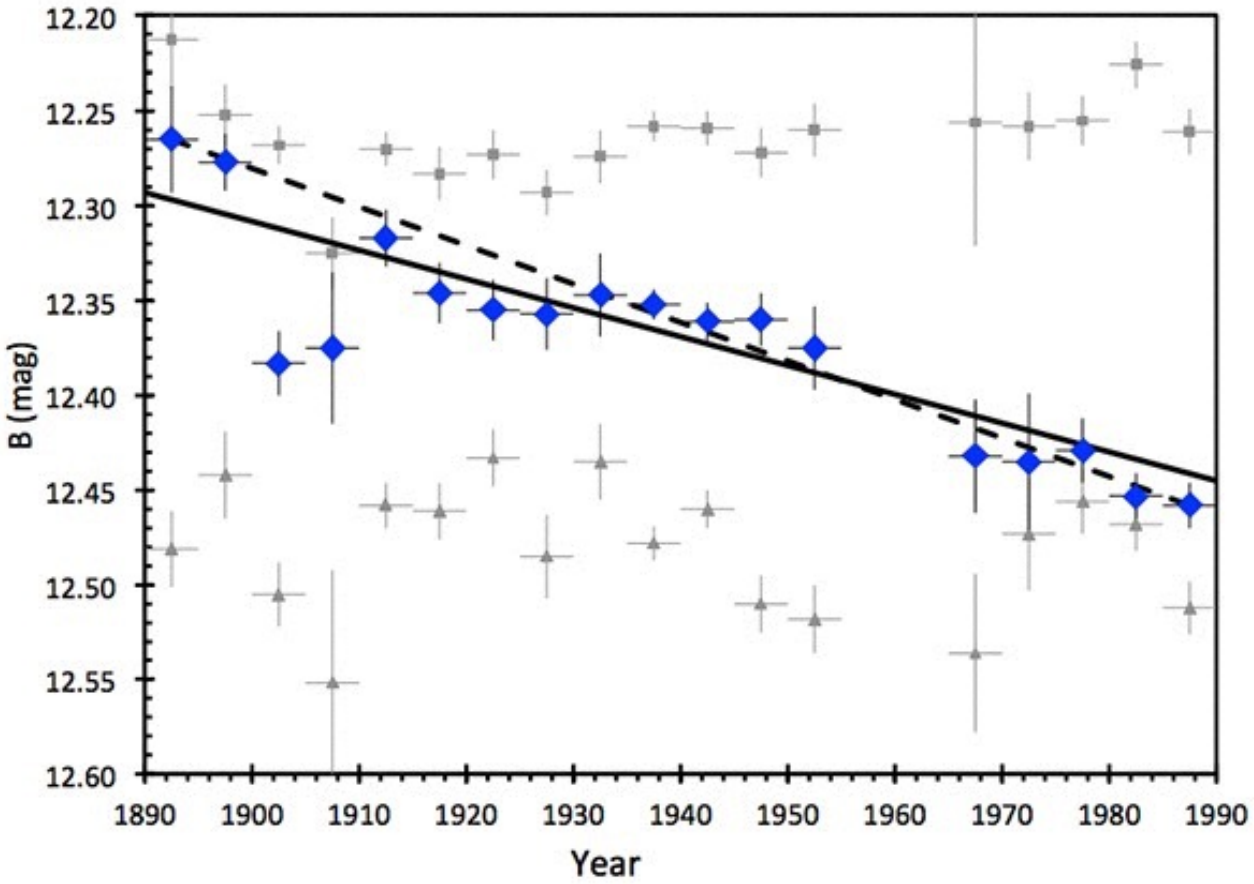
Feature Correlations



KIC 8462852 (aka Tabby's star aka WTF star)



nir and UV flux



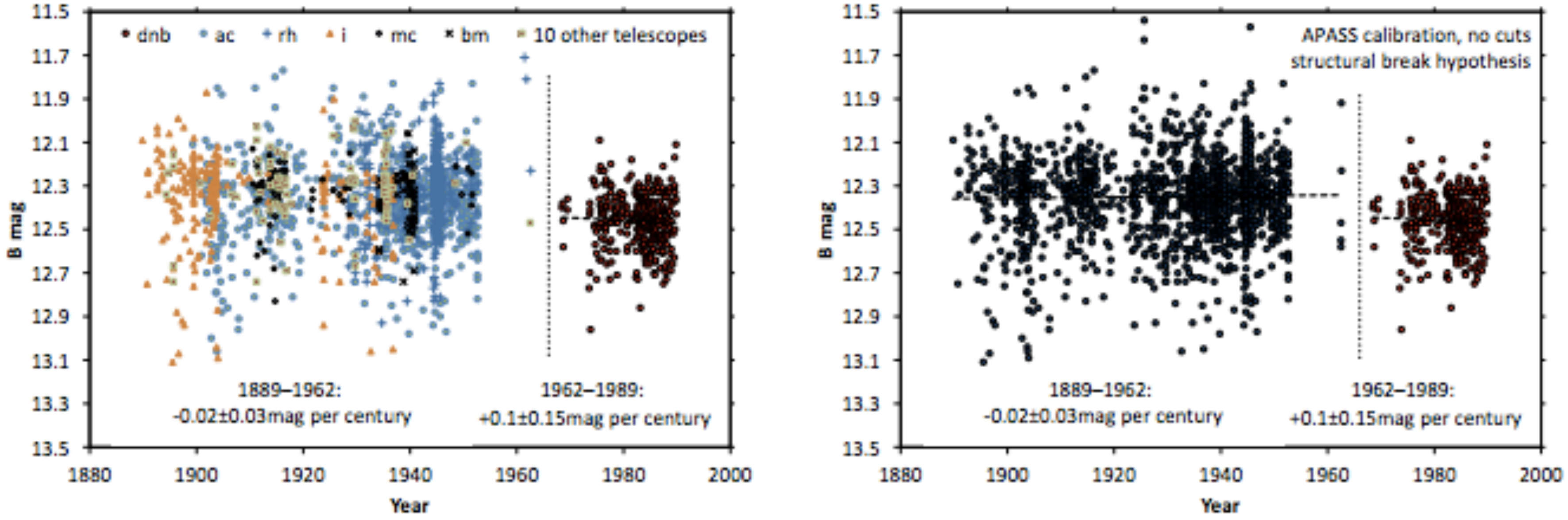
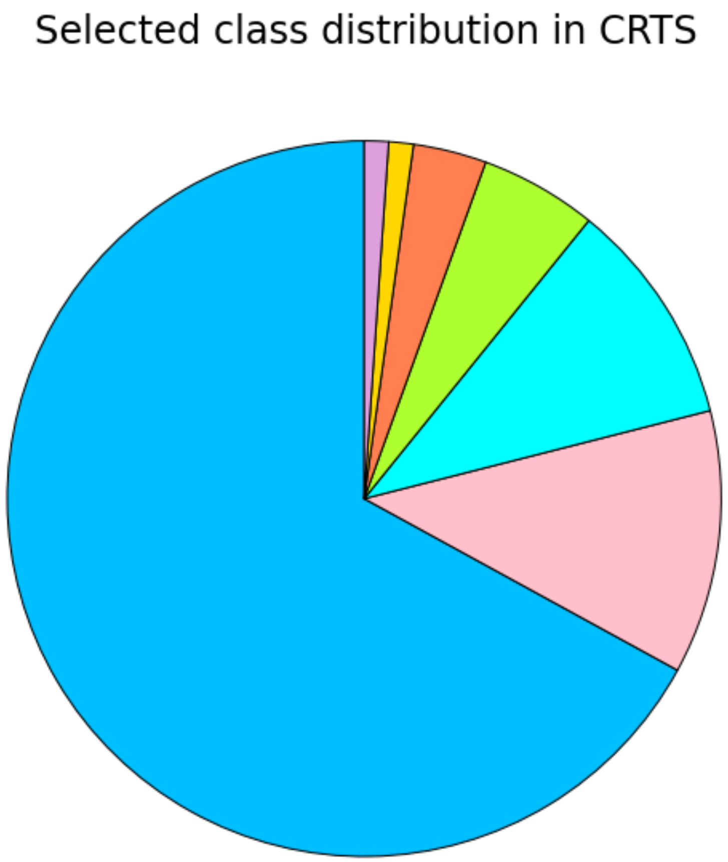
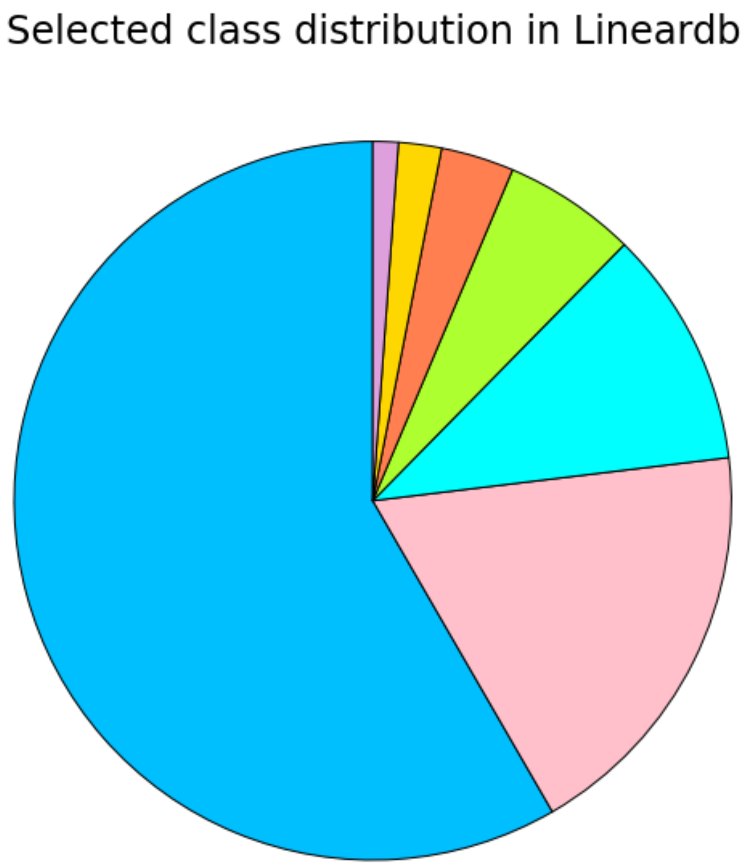


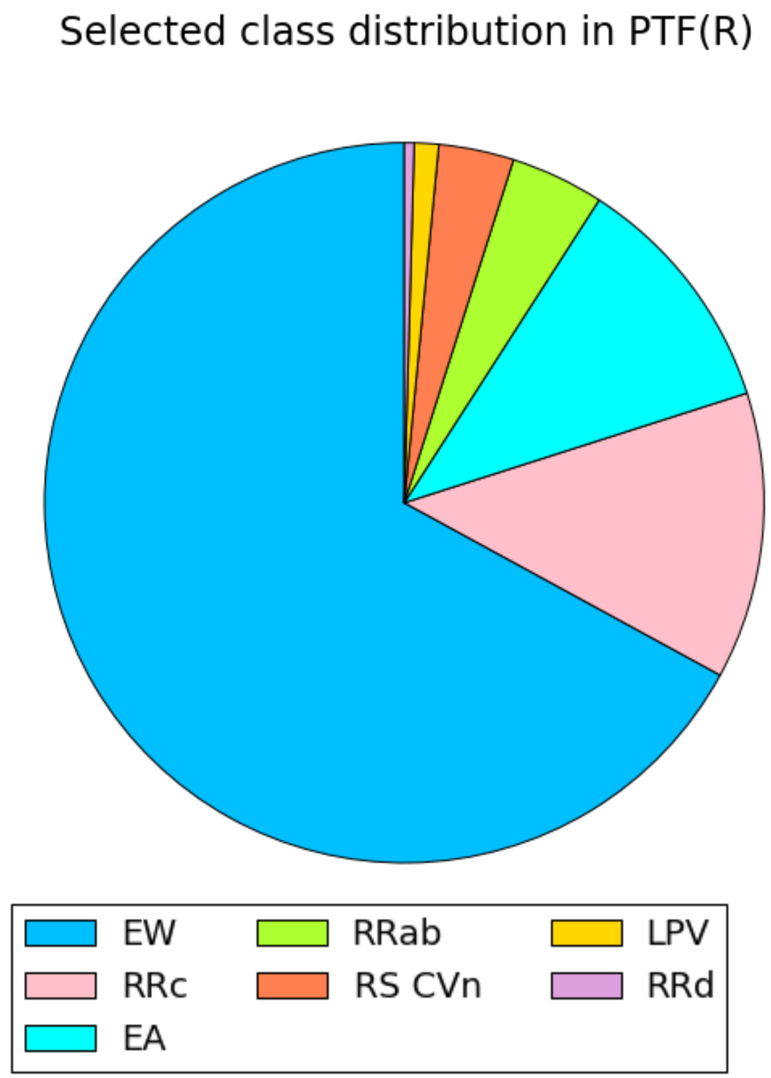
FIG. 3.— Hypothesis of a structural break. Left: The data before 1962 comes from 16 different telescopes, while the data after 1962 (red symbols) comes from only one telescope and shows an offset. Right: Linear regressions for both segments separately indicate constant luminosity within the errors. We hypothesize that the structural break is due to a different technology used after 1962 in "dnb" data, e.g. due to a different emulsion.

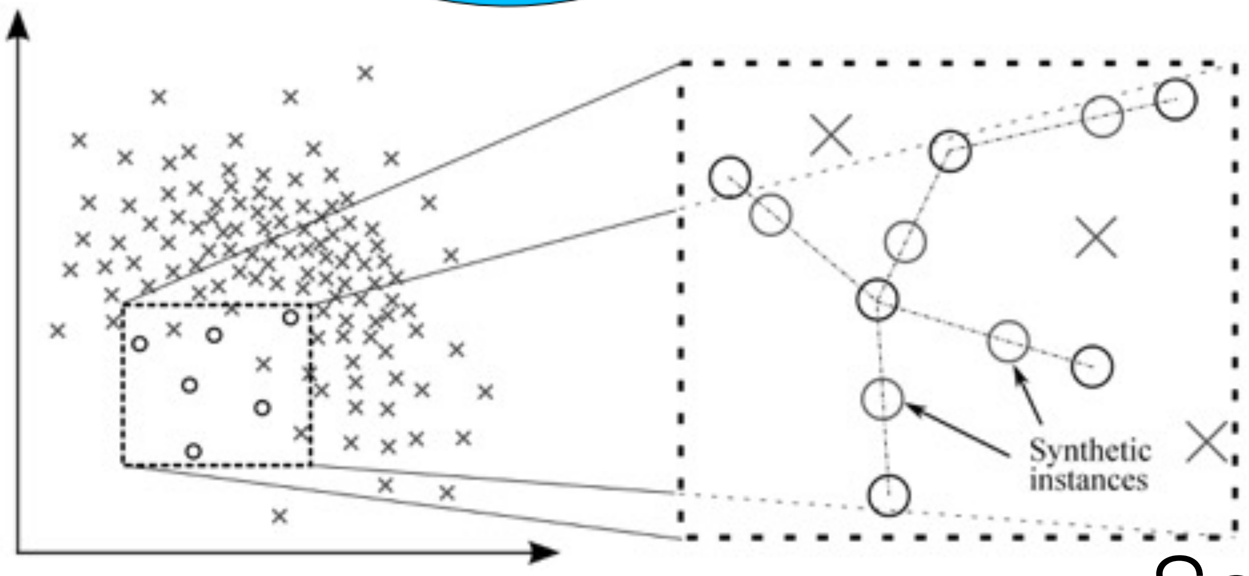
Fading or not fading?

50K Variables from CRTS   Drake et al. 2014



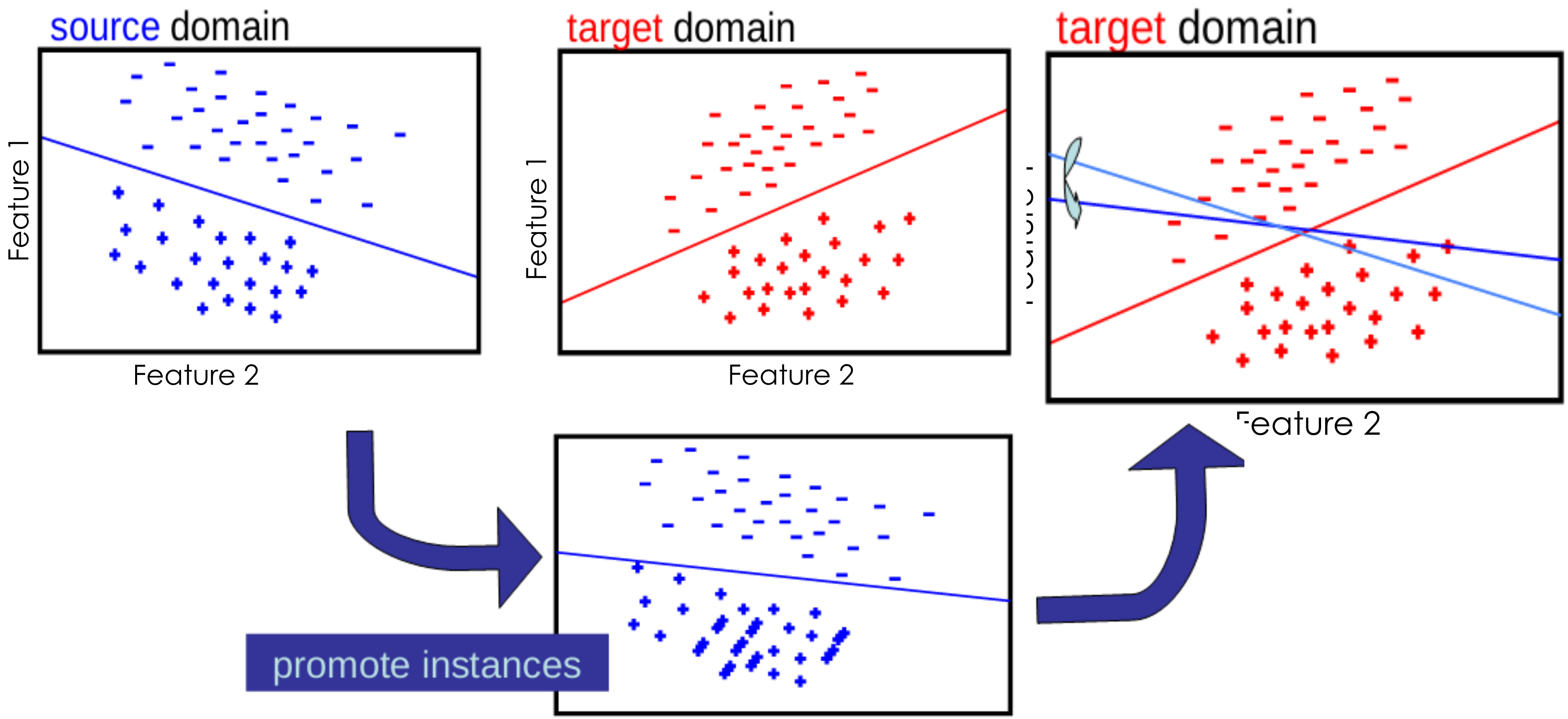






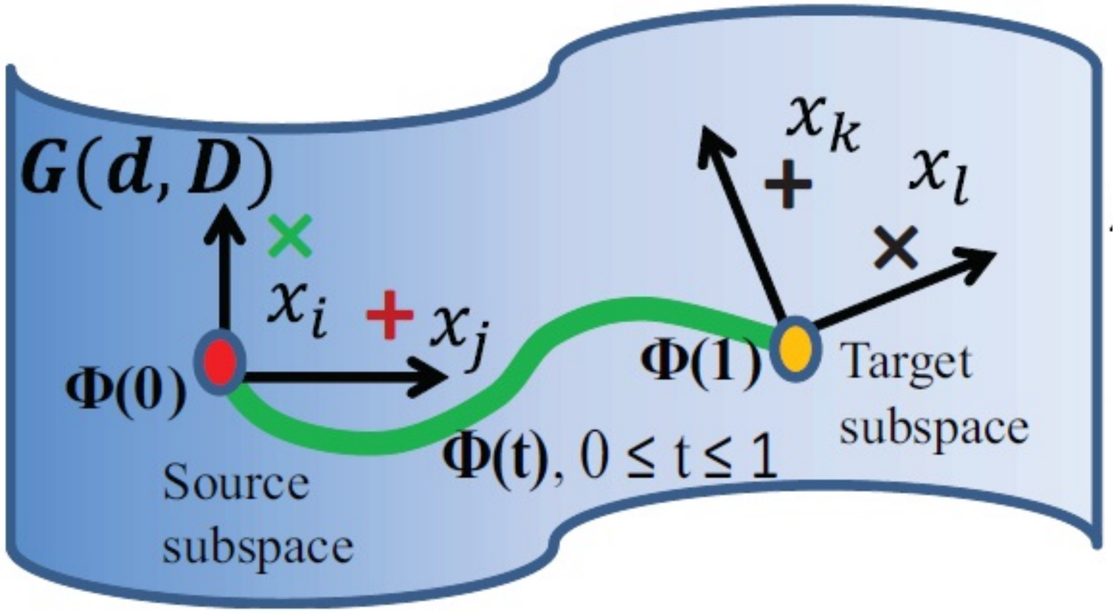
SMOTE and
Sampling with replacement
used to take care of unbalancedness

If you had just two features



Geodesik Flow Kernel

- Integrate flow of subspace: S to T
- Kernel incapsulates incremental changes between subspaces
- Kernel converts domain specific features into invariant ones (Gong et al. 2012)

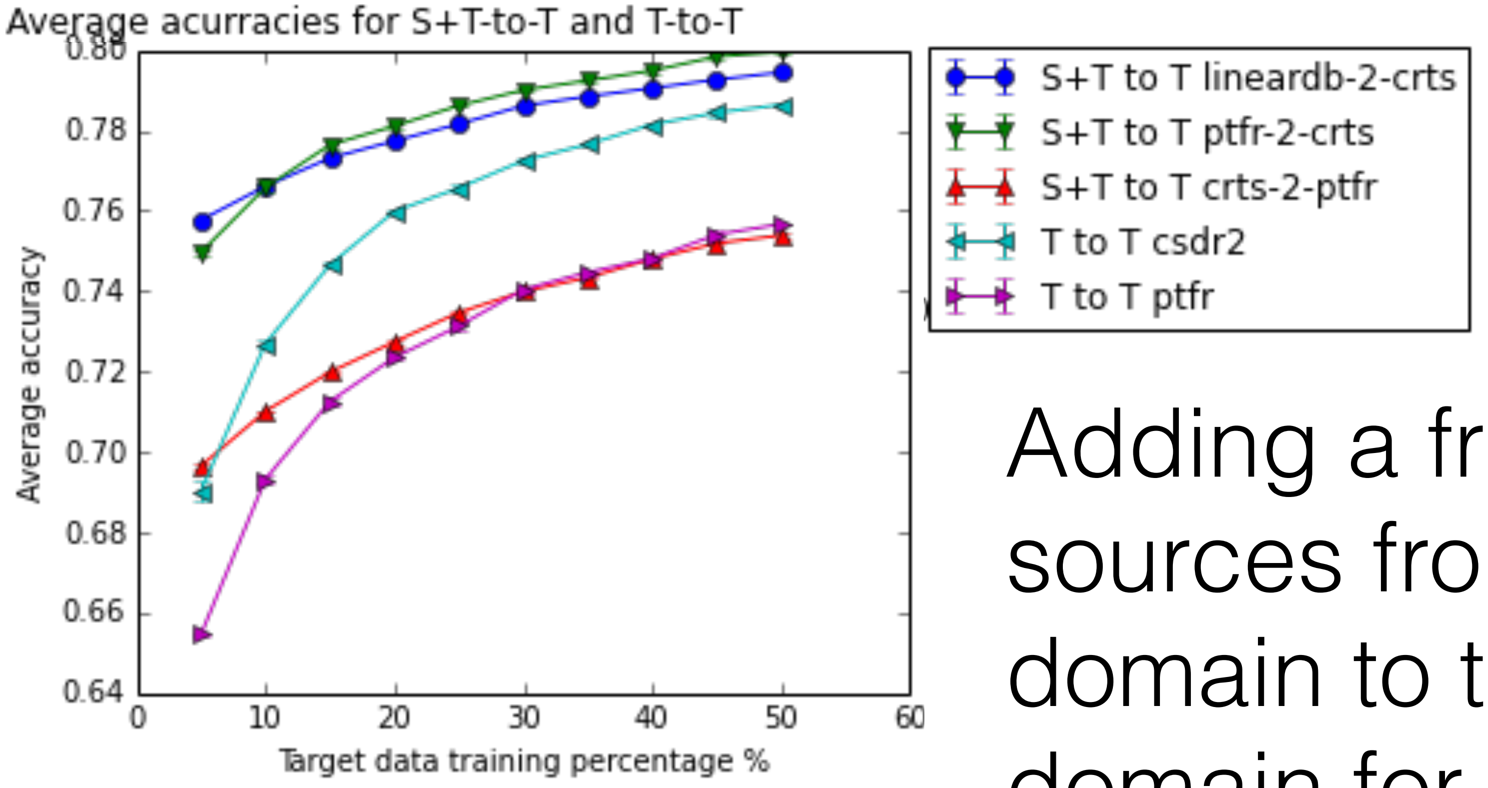


Co-Domain Adaptation

- Slow adaptation from S to T
- Add best target objects in each round
- Elect shared S and T subsets from training and unlabelled data (Chen et al. 2011)

$L = D_S \cup D_T^l$

$U = D_T^u$



Average Accuracy

Target Data Training %

Adding a fraction of sources from the target domain to the source domain for training improves performance

Summary of challenges

- **Characterize/Classify as much with as little data as possible**
- **Only a small fraction are rare - find/characterize them early**
- **A variety of parameters - choose judiciously**
- **Real-time computation is required - find ways to make that happen**
- **Metaclassification - combining diverse classifiers optimally**

Summary of challenges

- **Characterize/Classify as much with as little data as possible**
- **Only a small fraction are rare - find/characterize them early**
- **A variety of parameters - choose judiciously**
- **Real-time computation is required - find ways to make that happen**
- **Metaclassification - combining diverse classifiers optimally**

These challenges involve:

- Making sense of unparalleled volumes of structured and unstructured data in real-time, and
- Teaching machines how humans think by understanding pattern recognition when handling diverse types of data sources

Summary of challenges

- **Characterize/Classify as much with as little data as possible**
- **Only a small fraction are rare - find/characterize them early**
- **A variety of parameters - choose judiciously**
- **Real-time computation is required - find ways to make that happen**
- **Metaclassification - combining diverse classifiers optimally**

These challenges involve:

- Making sense of unparalleled volumes of structured and unstructured data in real-time, and
- Teaching machines how humans think by understanding pattern recognition when handling diverse types of data sources

Better tools to make sense of very sparse data and Streamlined workflows