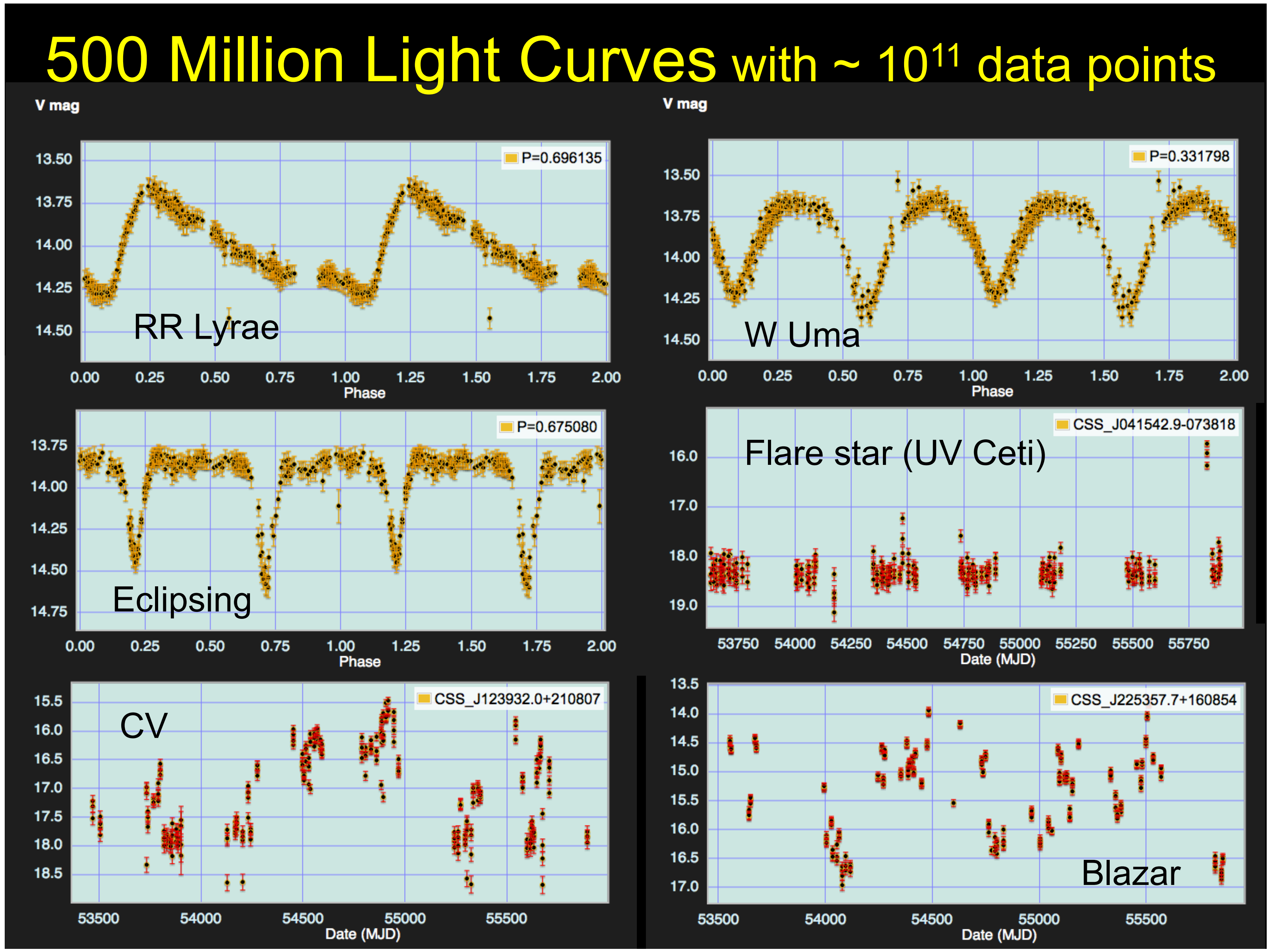
Irregular time series from wide-field surveys



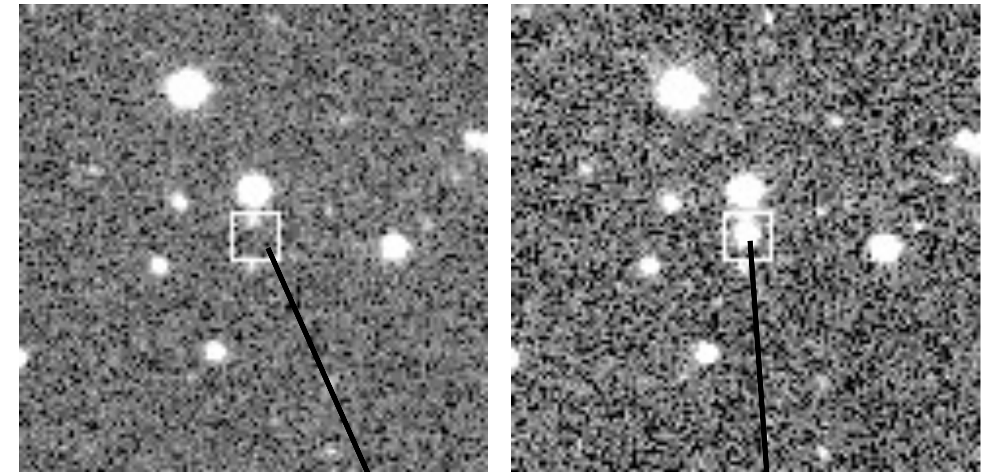
Ashish Mahabal
Astronomy and Center for Data Driven Discovery, Caltech

AY 119, 13 May 2019

Outline

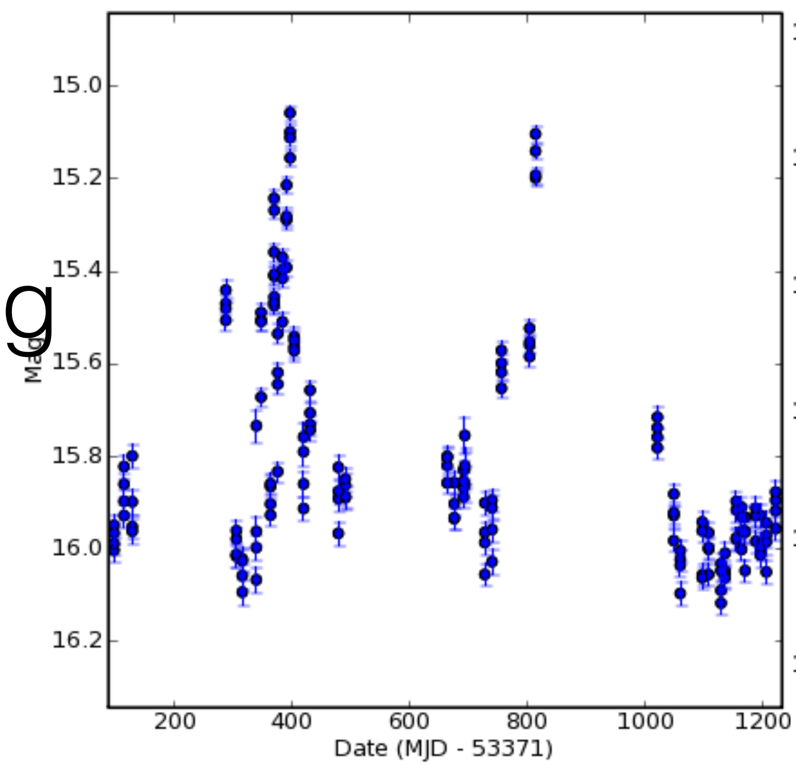
- Astronomical time series
  - their nuances
- examples from CRTS, ZTF etc.
- Statistical features and priors
- Period fitting
- reformatting, stochastic time series etc.
- [Examples/Exercises]

Time Series aka light-curves
we will encounter

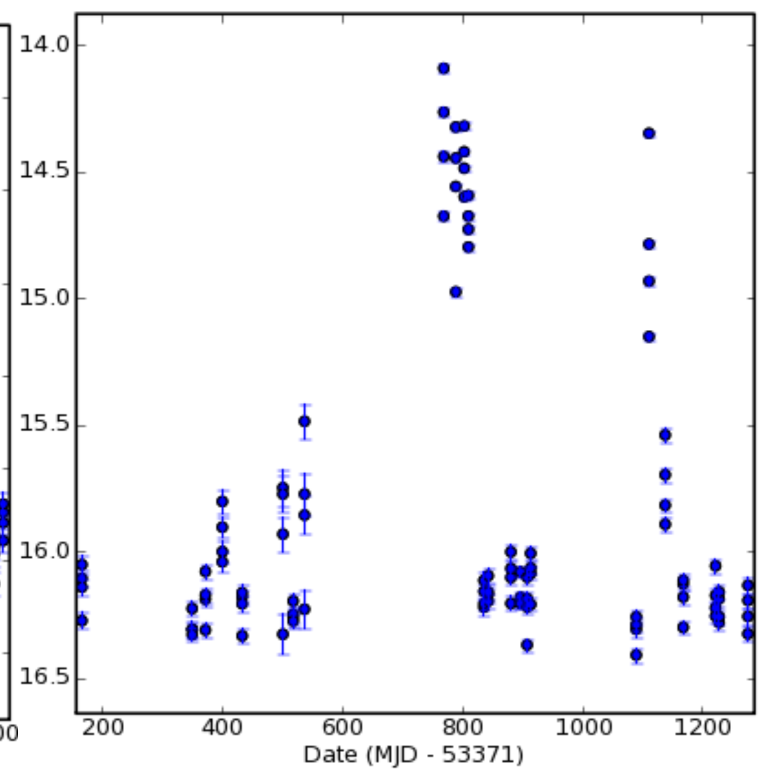


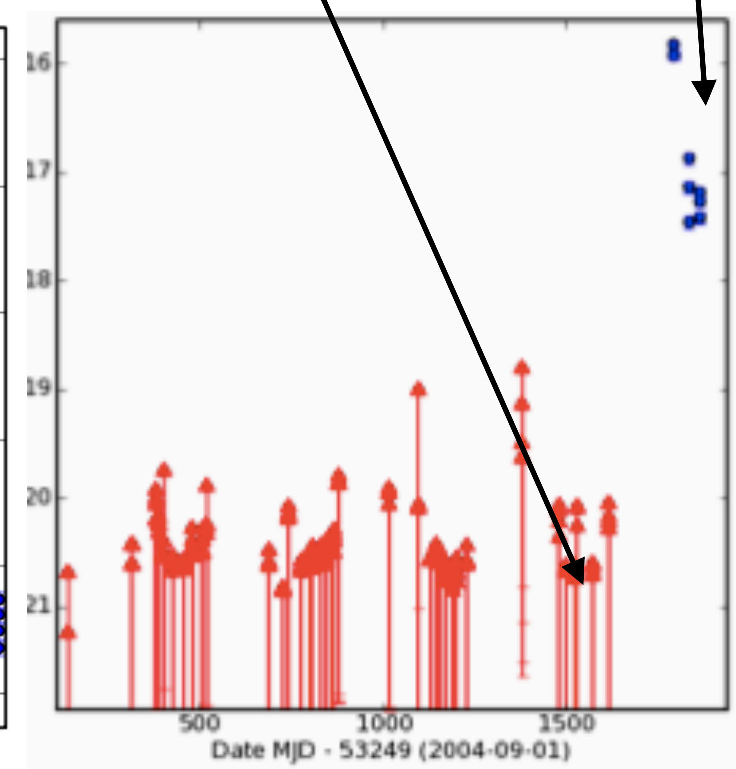
Blazar PKS0823+033

mag



CV 111545+425822

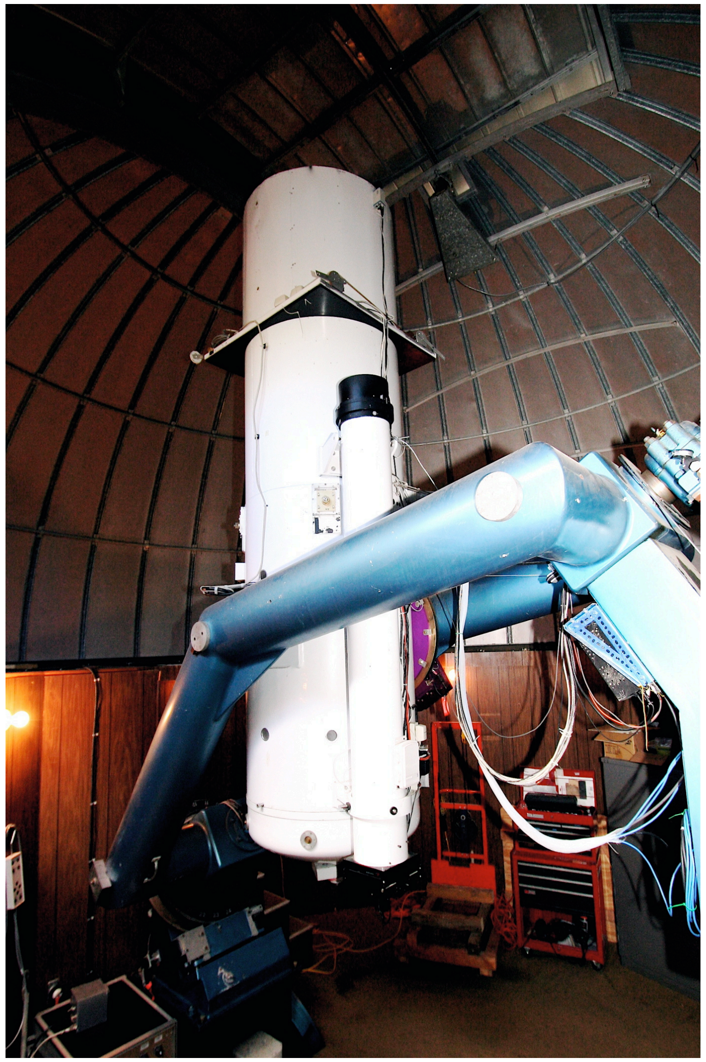




Supernova

Time (1000+ days)

magnitude is logarithmic, inversely scaled (flux)



CRTS

Transient
Searches

1m class telescopes
~20 mag

23000 sq. deg
(moon ~ 0.25 sq deg

hundreds of pointings
30 seconds each

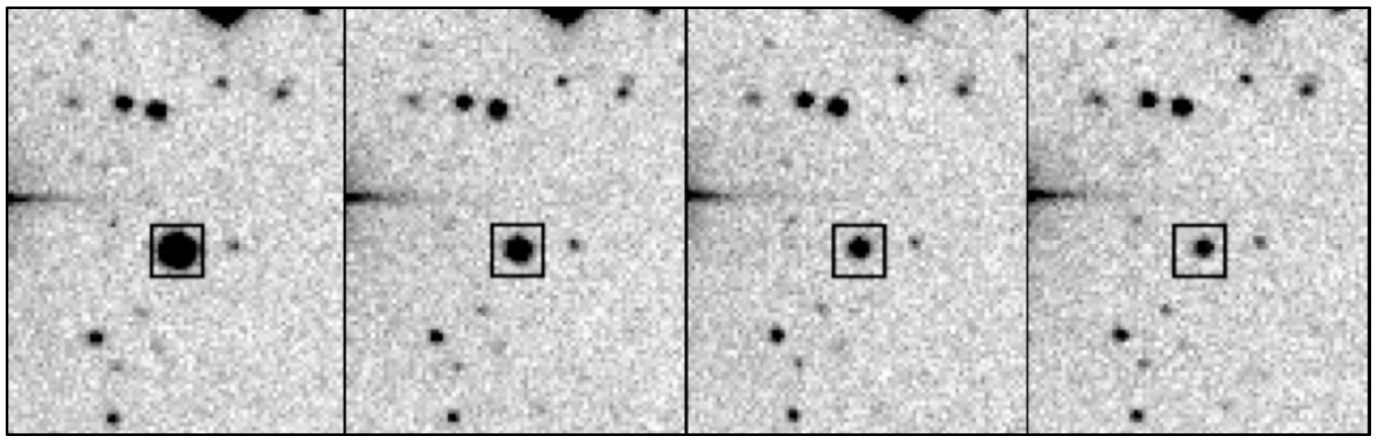
Open filter
~17 years
500M light-curves

DR1 MSIP (public) data

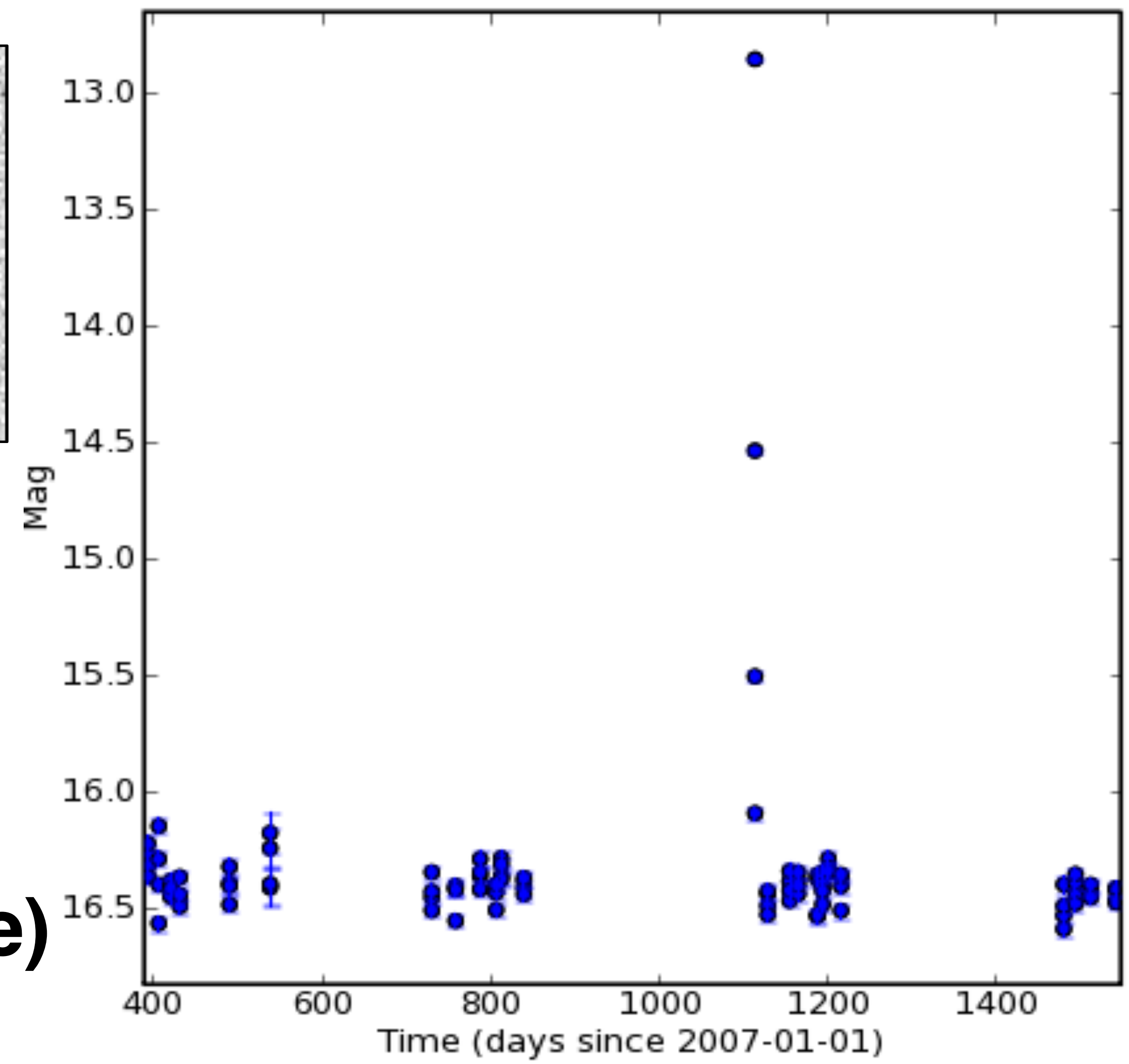
ztf.caltech.edu

Filter(s)	#lightcurves with $N_{obs} \geq 2$	#lightcurves with $N_{obs} \geq 5$	#lightcurves with $N_{obs} \geq 10$	#lightcurves with $N_{obs} \geq 20$
g	704,000,504	589,547,084	508,917,850	391,041,883
r	1,334,687,993	1,142,671,302	1,013,283,728	852,773,692
$g + r$	2,038,688,497	1,732,218,386	1,522,201,578	1,243,815,575

What is a transient?



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



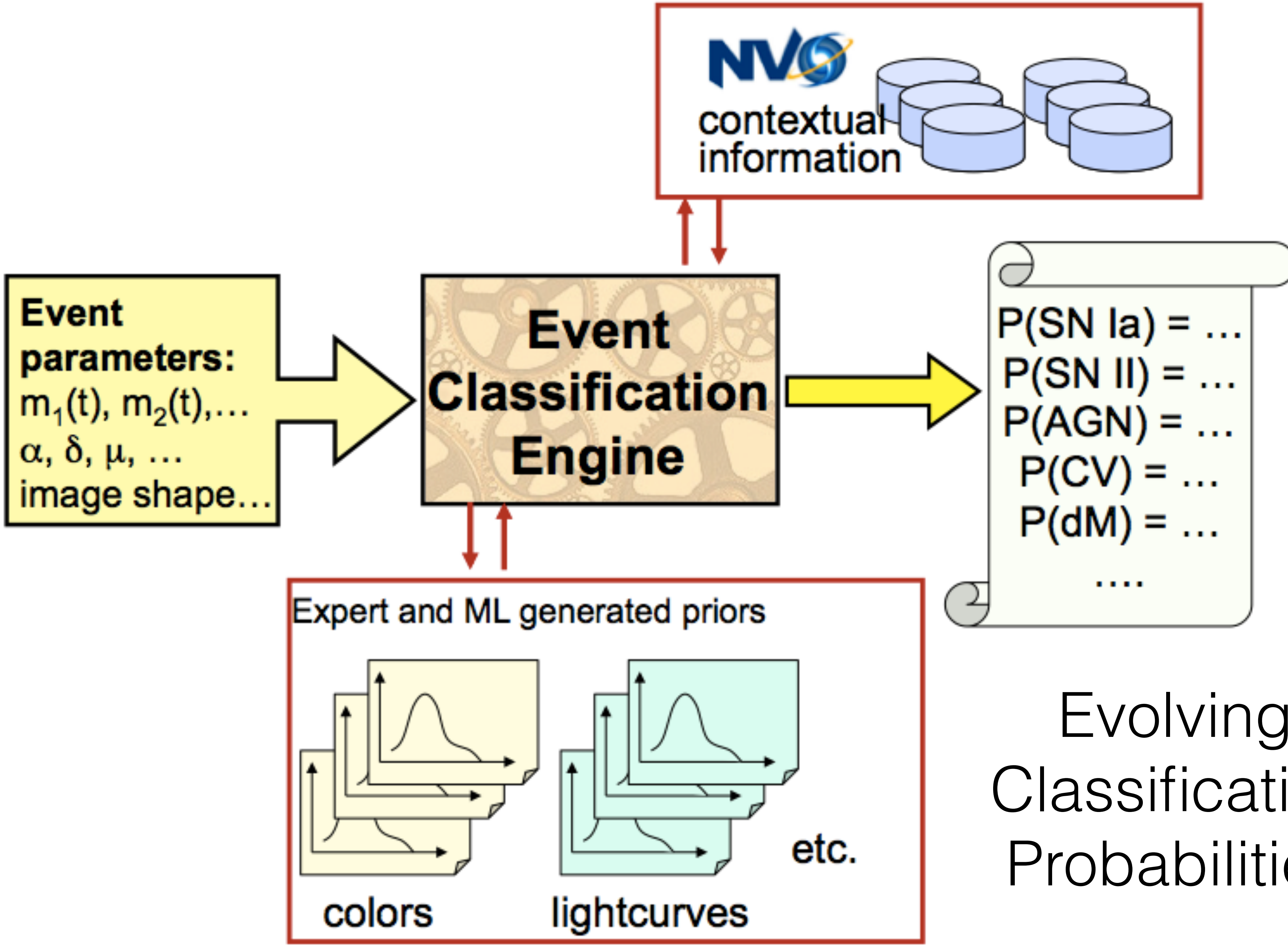
One that has a large brightness change (delta-magnitude) within a short timespan (small delta-time)

light-curve

transient | variable?

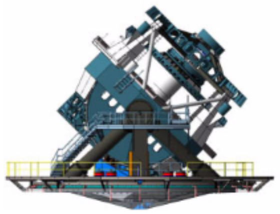
real-time | archival?

A few years ago …



Evolving Classification Probabilities

ZTF is doing a good fraction of this today!



LSST data volume and scientific yields

Raw Data

- Two 6.4-gigabyte images (one visit) every 39 seconds (15TB per night)
- ~1000 visits each night, ~300 nights a year
- Up to 450 calibration exposures per day

Level 1

- Can detect >10 million real time events per night, for 10 years
- Changes detected, transmitted, within 60 seconds of the observation

Level 2

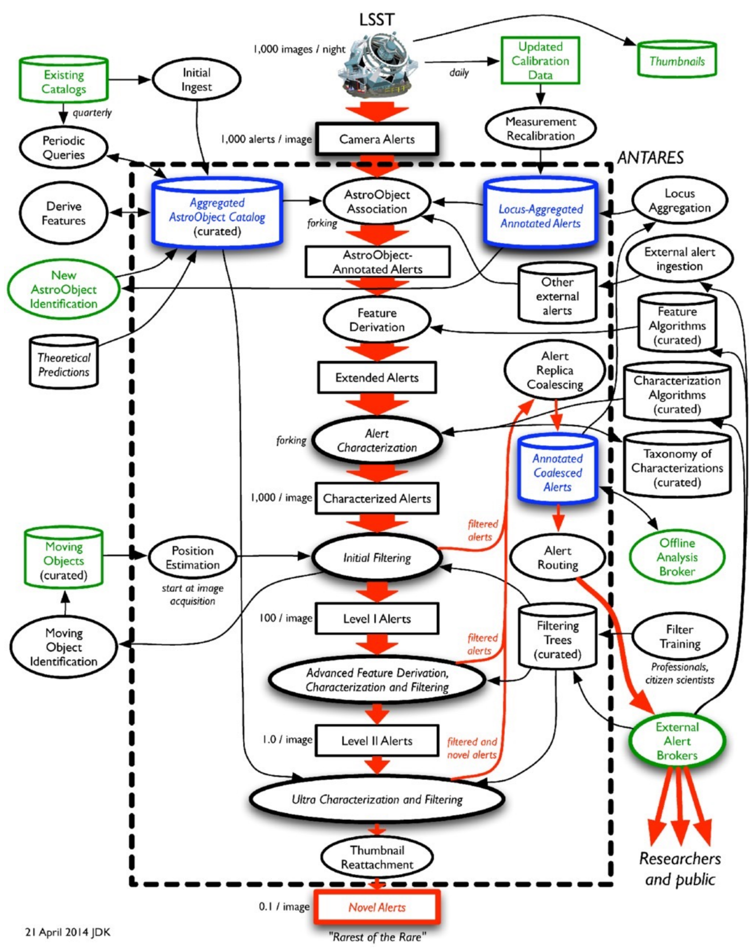
- Observe ~38 billion objects (24B galaxies, 14B stars)
- Collect ~5 trillion observations ("sources") and ~32 trillion measurements ("forced sources") in a 20 PB catalog

Level 3

- User databases and workspaces ("mydb")
- Making the LSST software available to end-users
- Feeding the data back to the community

Active learning to minimize follow-up

10^7 transients



10^3 rare transients

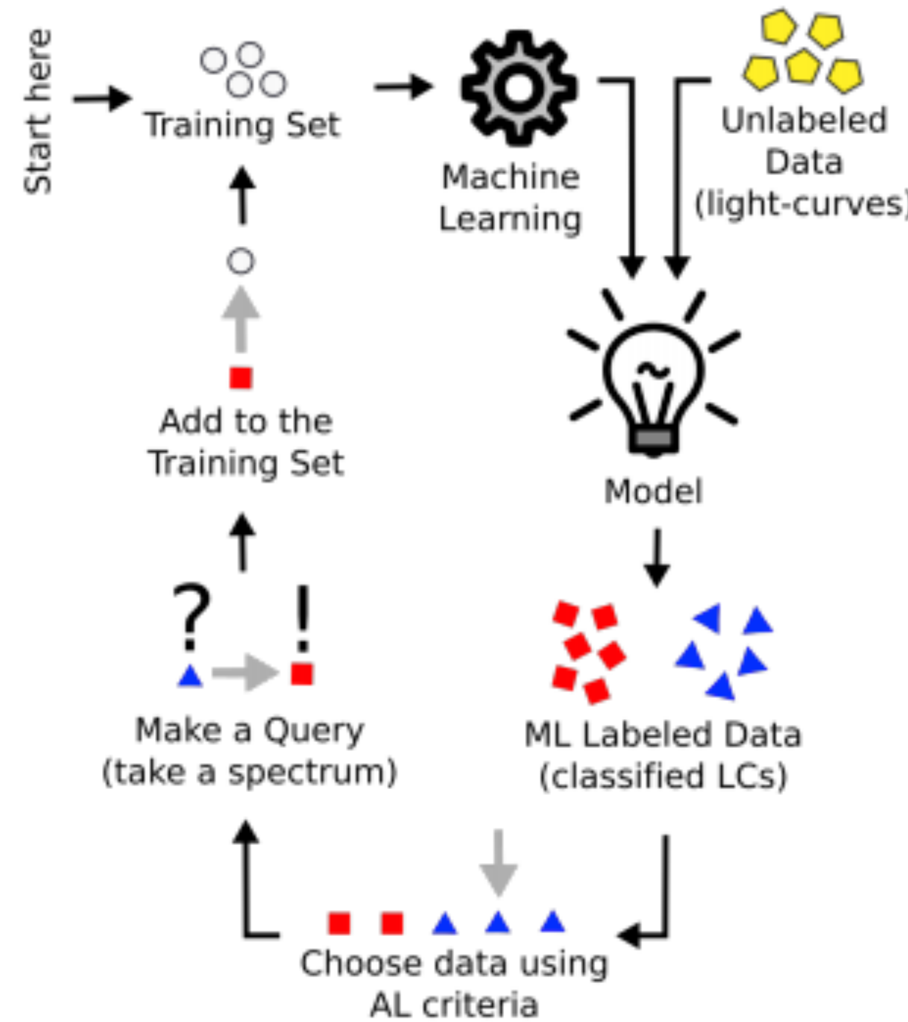
Connecting to Brokers

ALeRCE.science

AMPEL

Antares

Lasair



Ishida et al. 2018, arXiv:1804.03765

Accuracy of 80% reached in 100 days of observations, far above the canonical rate

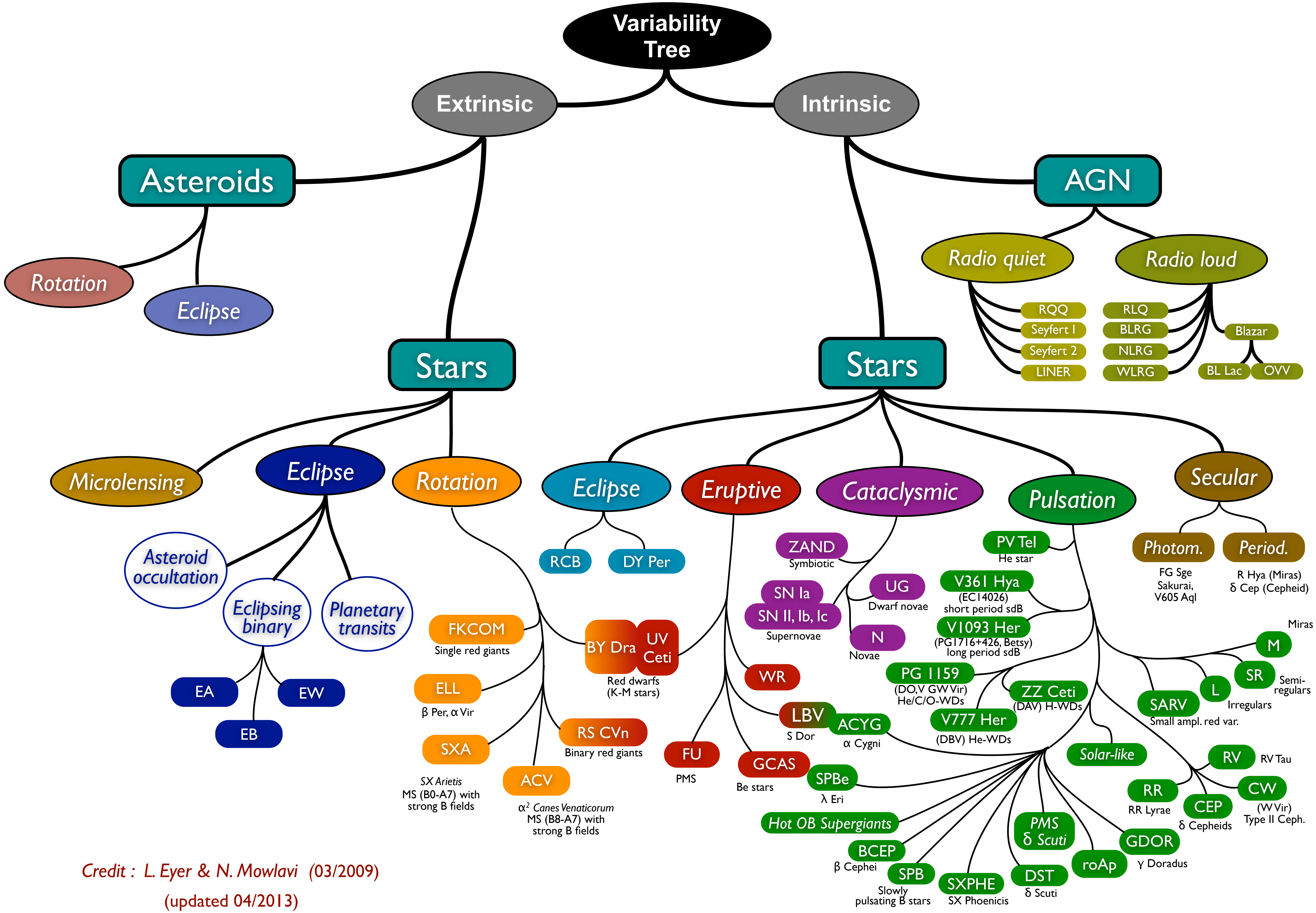
Background domain knowledge

Broad classes in astronomy

Aim:

- **Understanding the Universe**
    - **classification -> understanding**
- Solar System - moving objects
- Stars in our Galaxy - variables, proper motion
- Extragalactic - mostly transient

Variability tree: Many nodes have further subdivisions



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

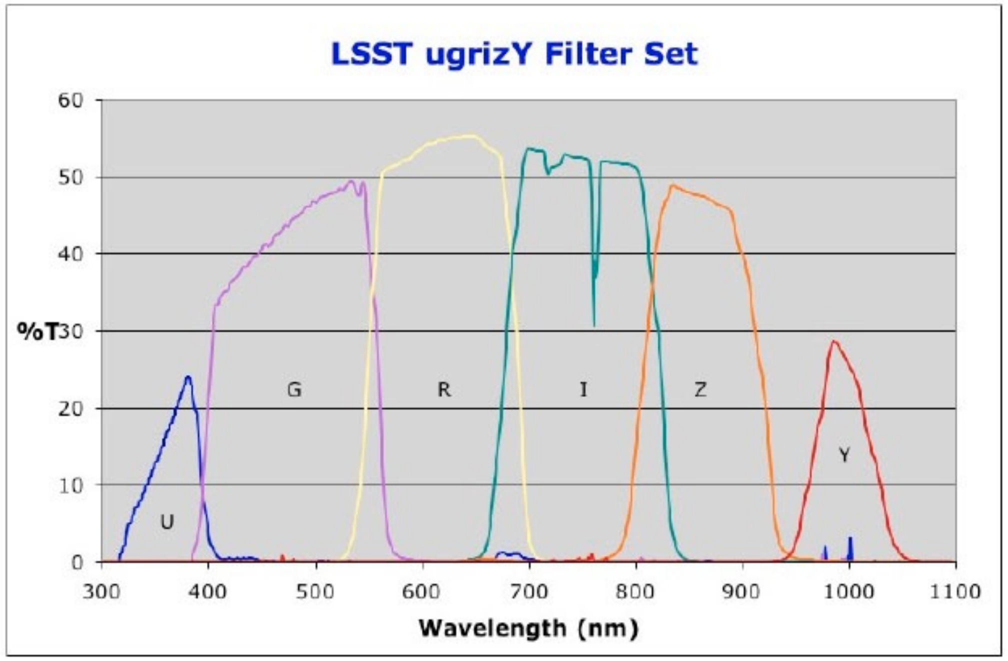
Slide from Lucianne Walkowicz

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

Broad classes in astronomy



- Solar System - moving objects
- Stars in our Galaxy - variables, proper motion
- Extragalactic - mostly transient

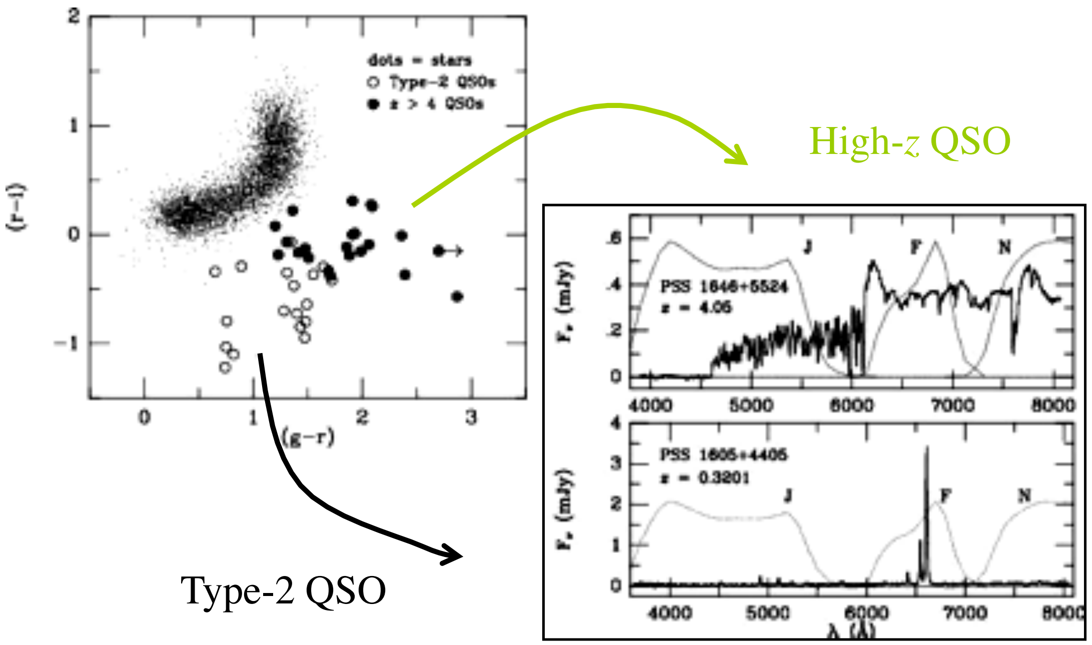
Aim:

- Understanding the Universe
  - classification -> understanding

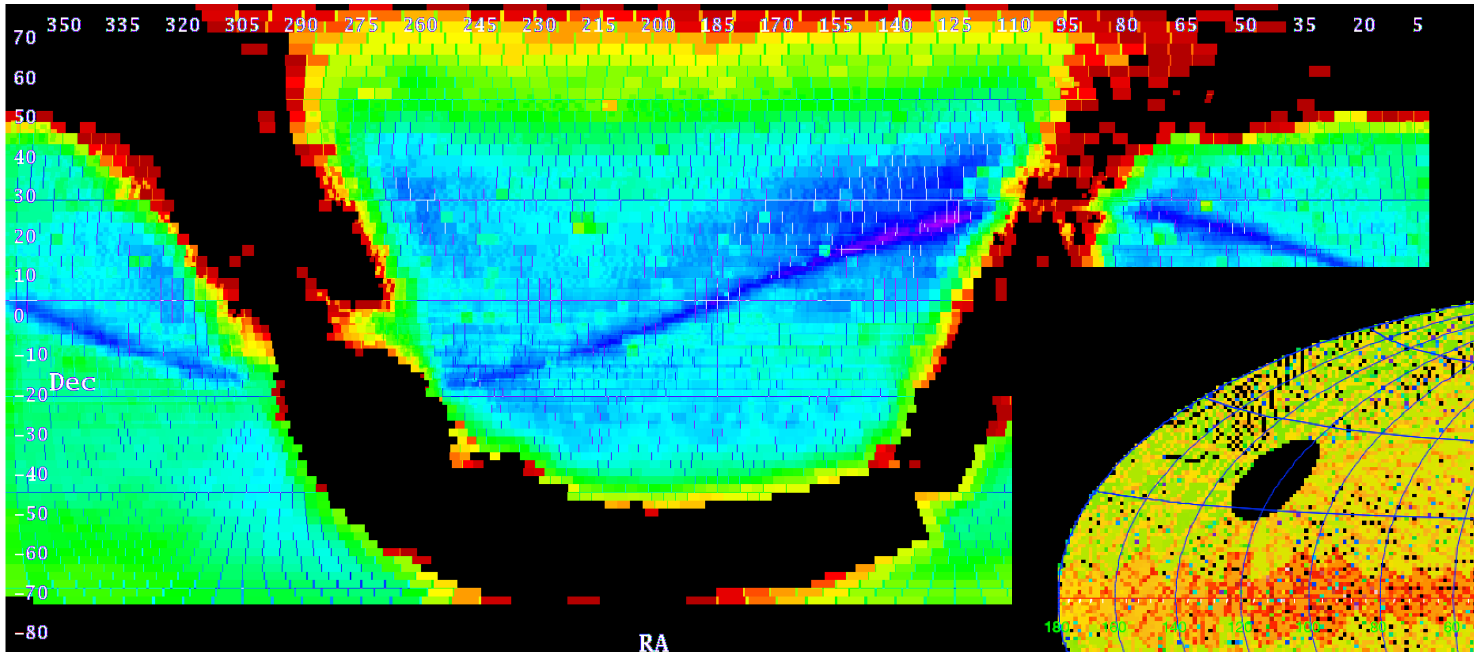
Our windows:

- **Spectroscopy: ideal - but expensive**
- **Colors and time-series**
  - **characterization**
  - **clustering (unsupervised)**
- **Other (Polarization, GW)**

An Example: Discoveries of High-Redshift Quasars and Type-2 Quasars

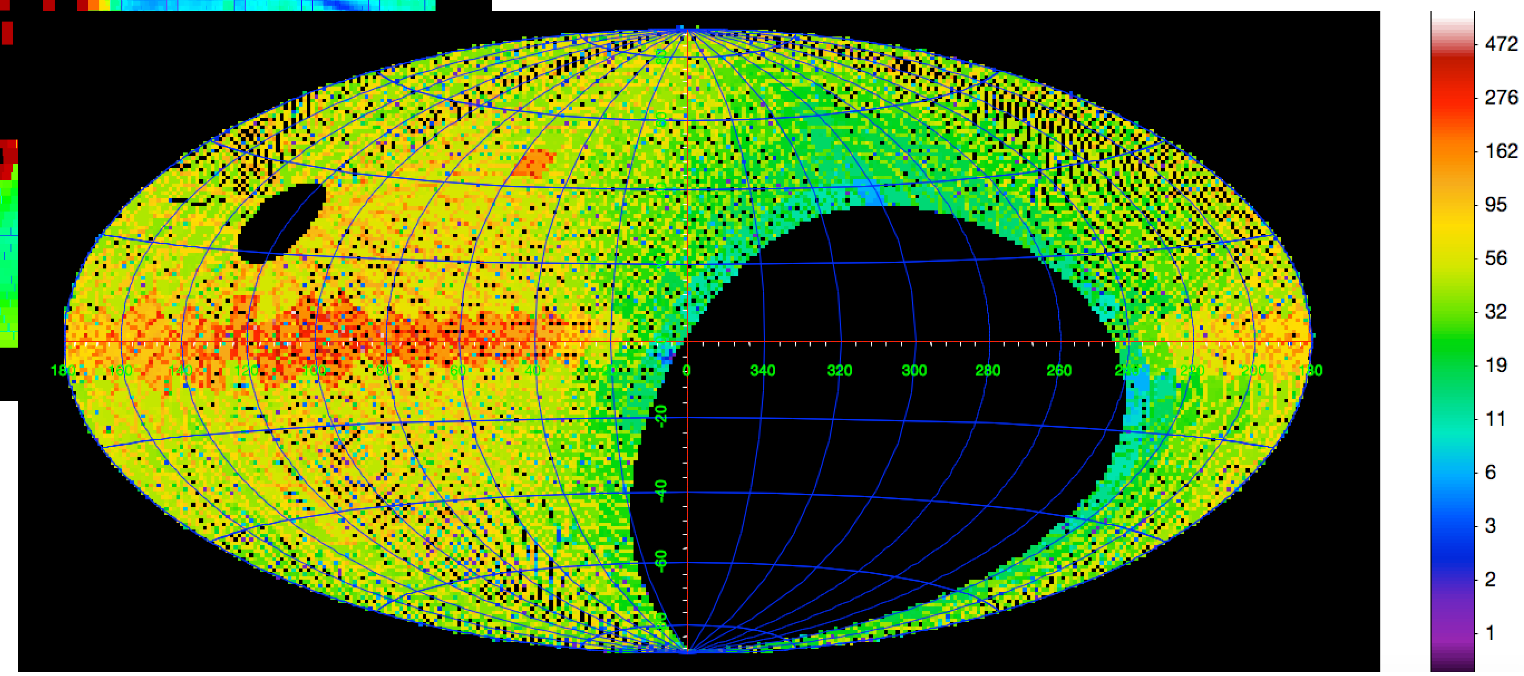


From snapshots to (slow) movies of the sky

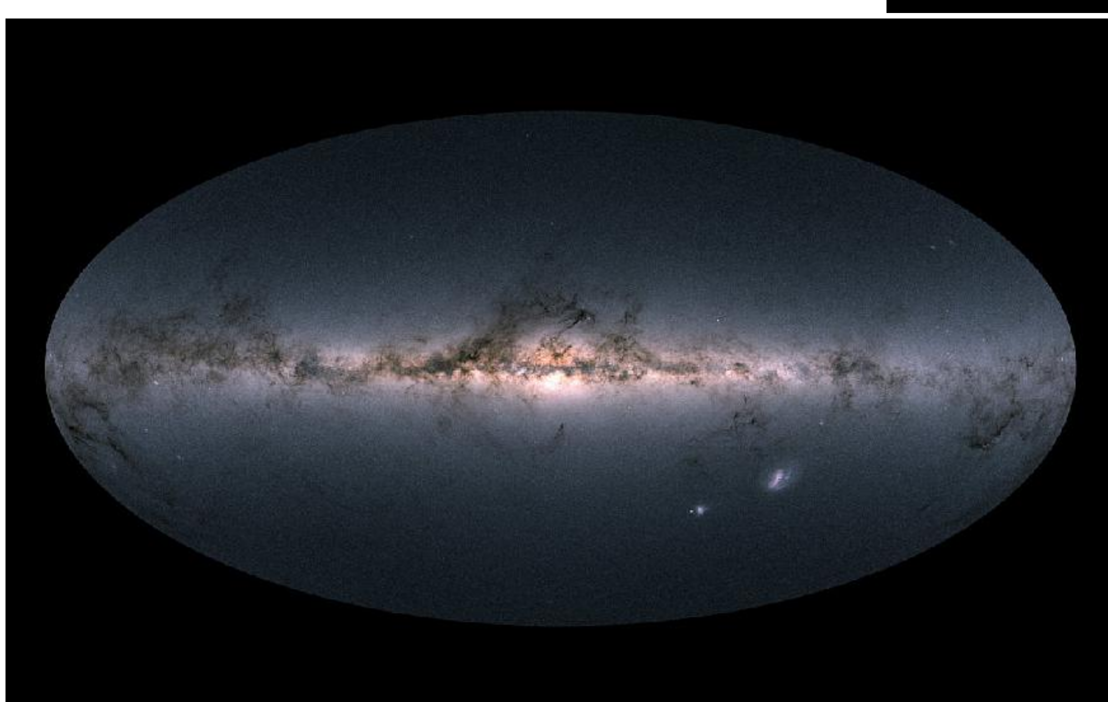


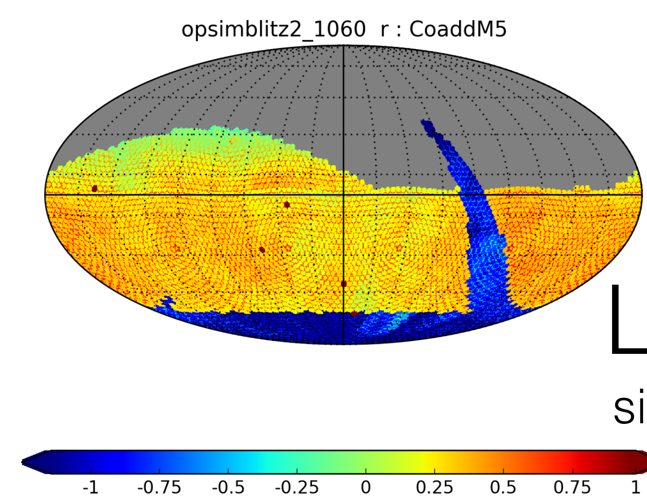
CRTS

ZTF



Gaia





LSST
simulated

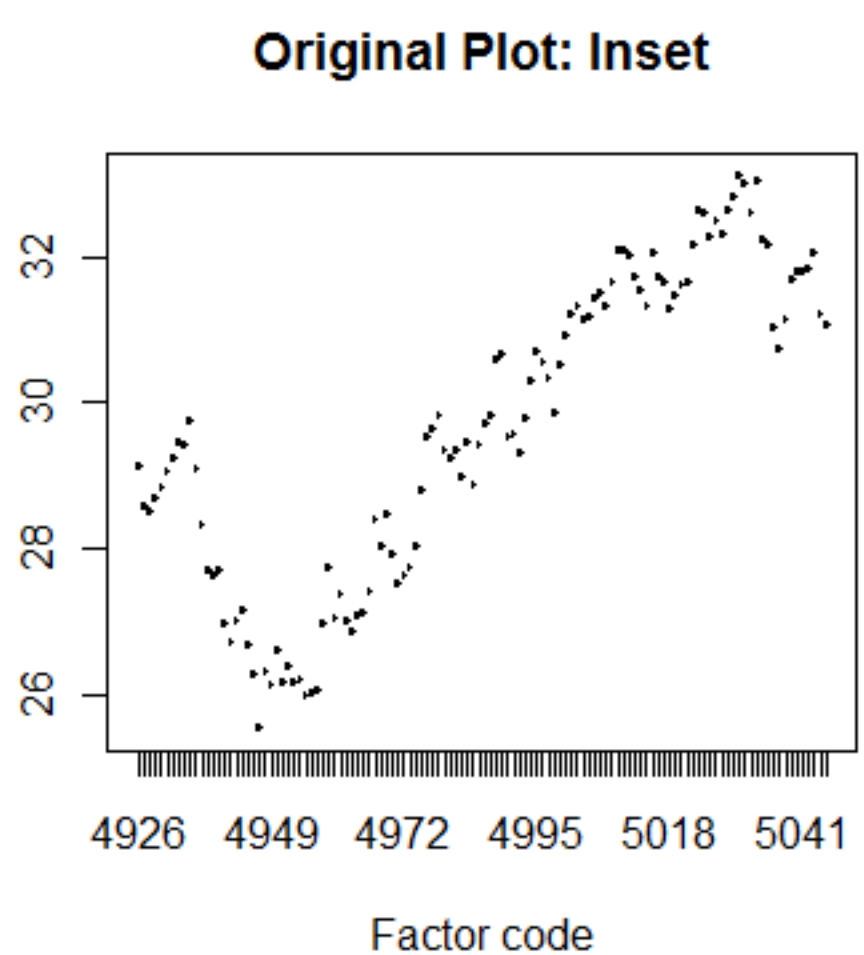
What do survey's do?

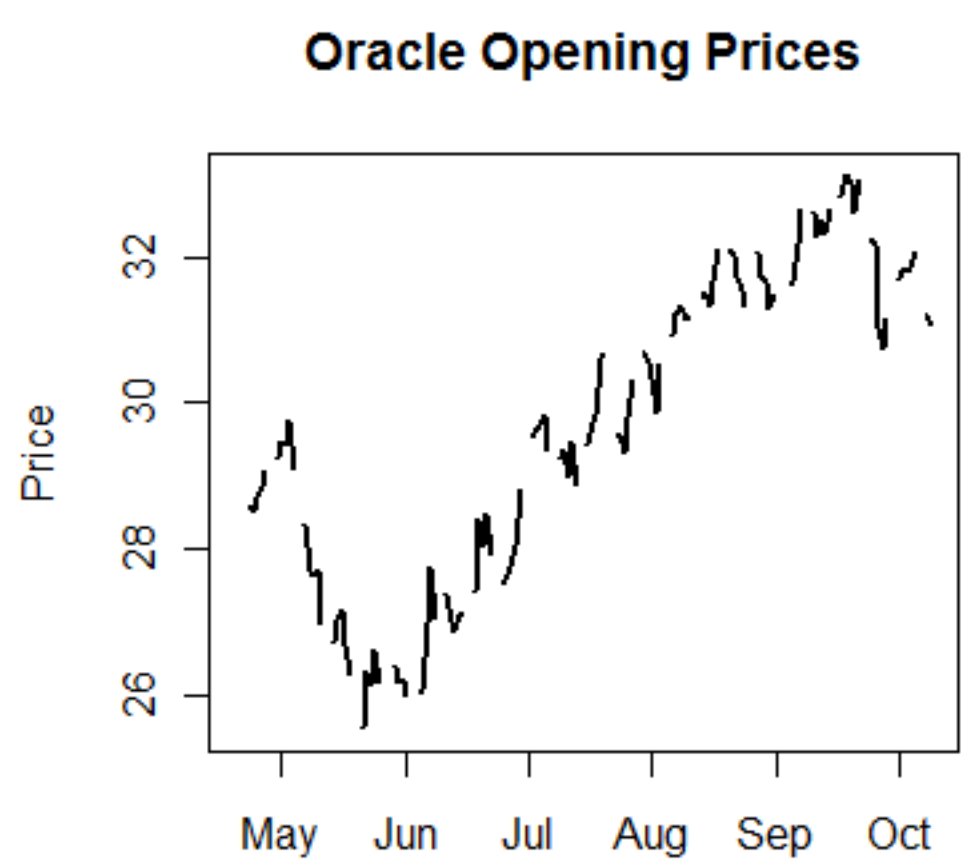

- Pick low-hanging fruit
  - select best objects, easy science
  - get spectroscopy
- That does push the envelope
  - but also leaves gaps

1000 30-sec epochs
10 years
3*10^4/3*10^8
1mm in 10m

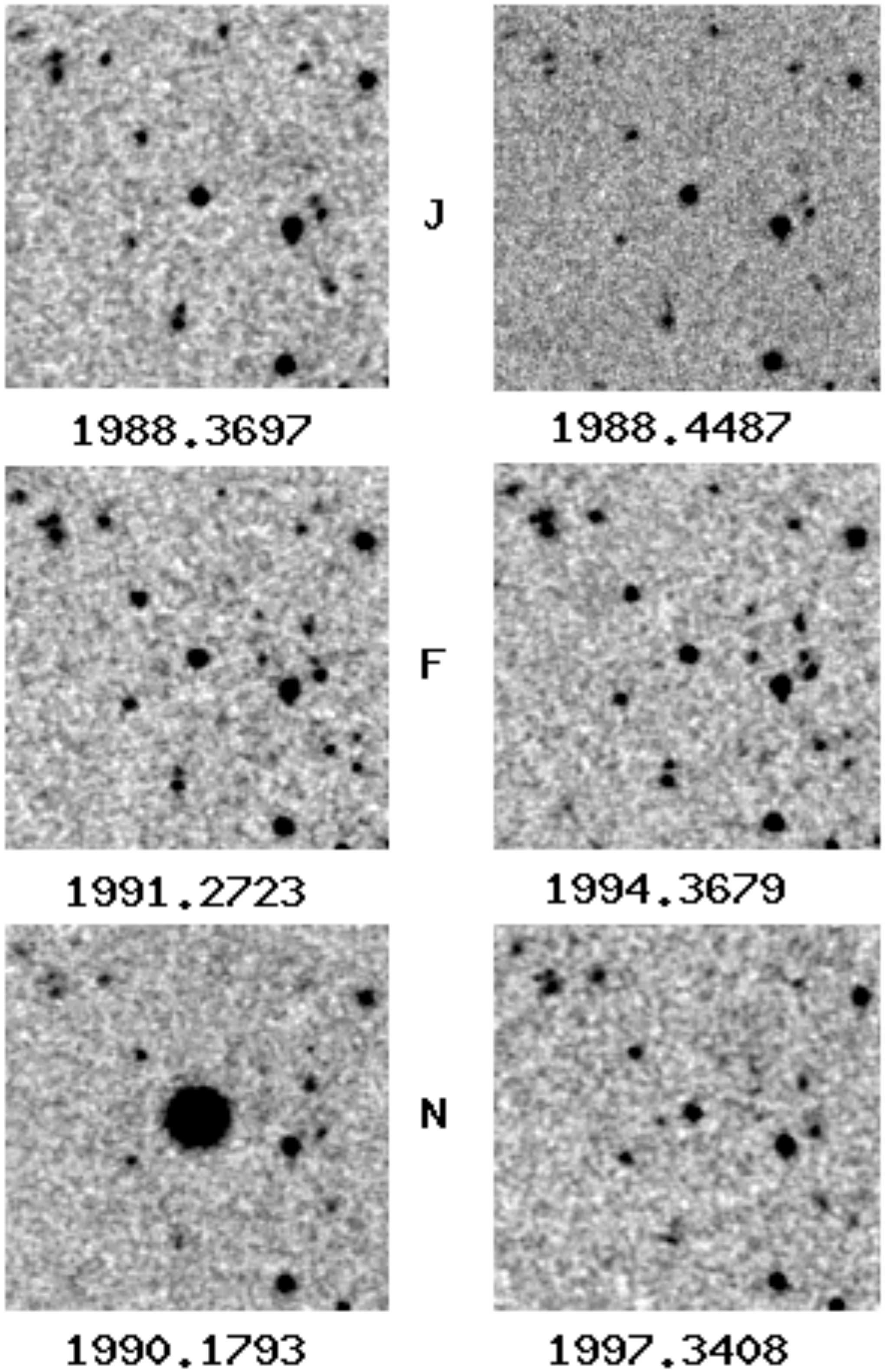
How gaps can be misleading

- Variations as a function of time
  - Financial
    - diurnal, regular, accurate, (almost) continuous



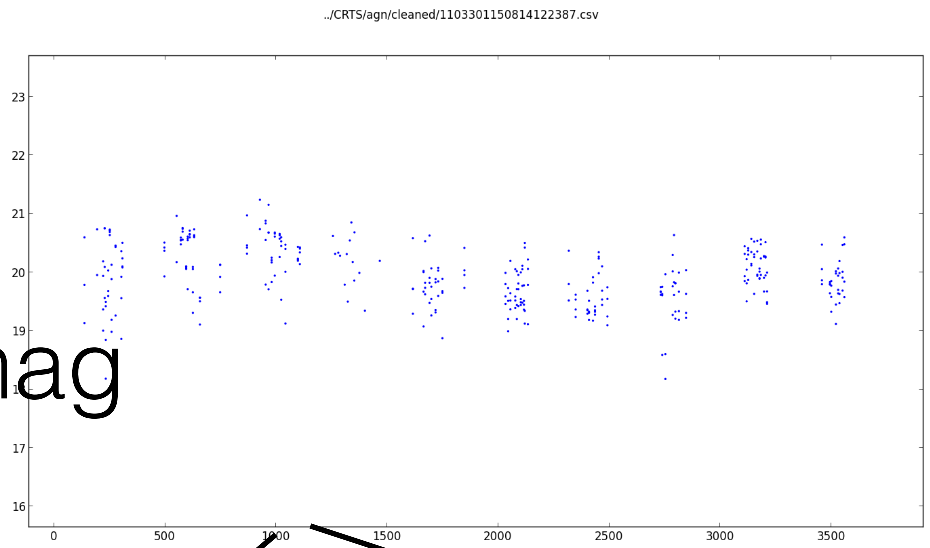


stackexchange



DPOSS: 3 filters
large area,
serendipitous overlap
Separation: 1 hr to 15 years

6.5 degree plates
centers separated by 5

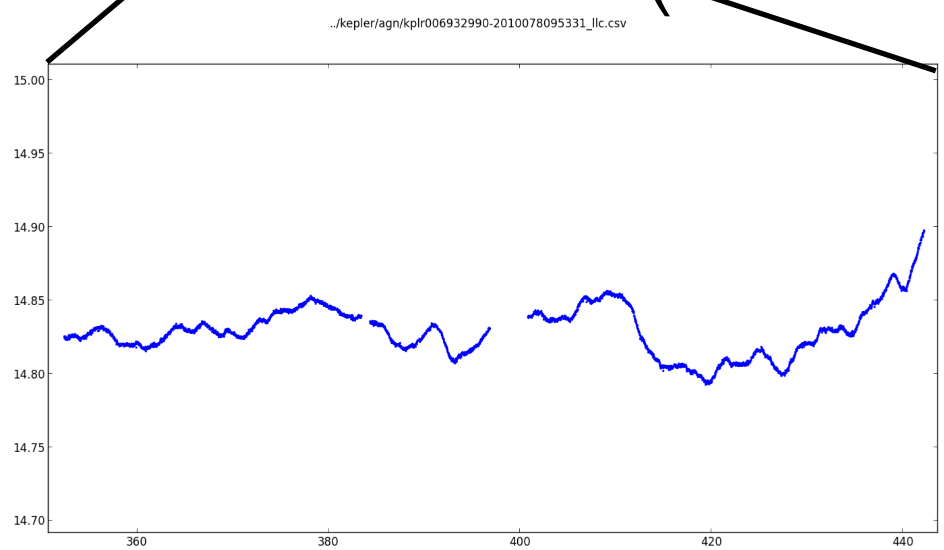


CRTS

Kepler - small area non-sparse

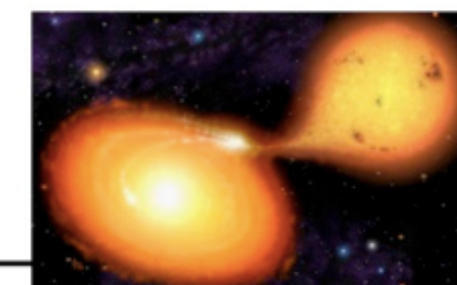
mag

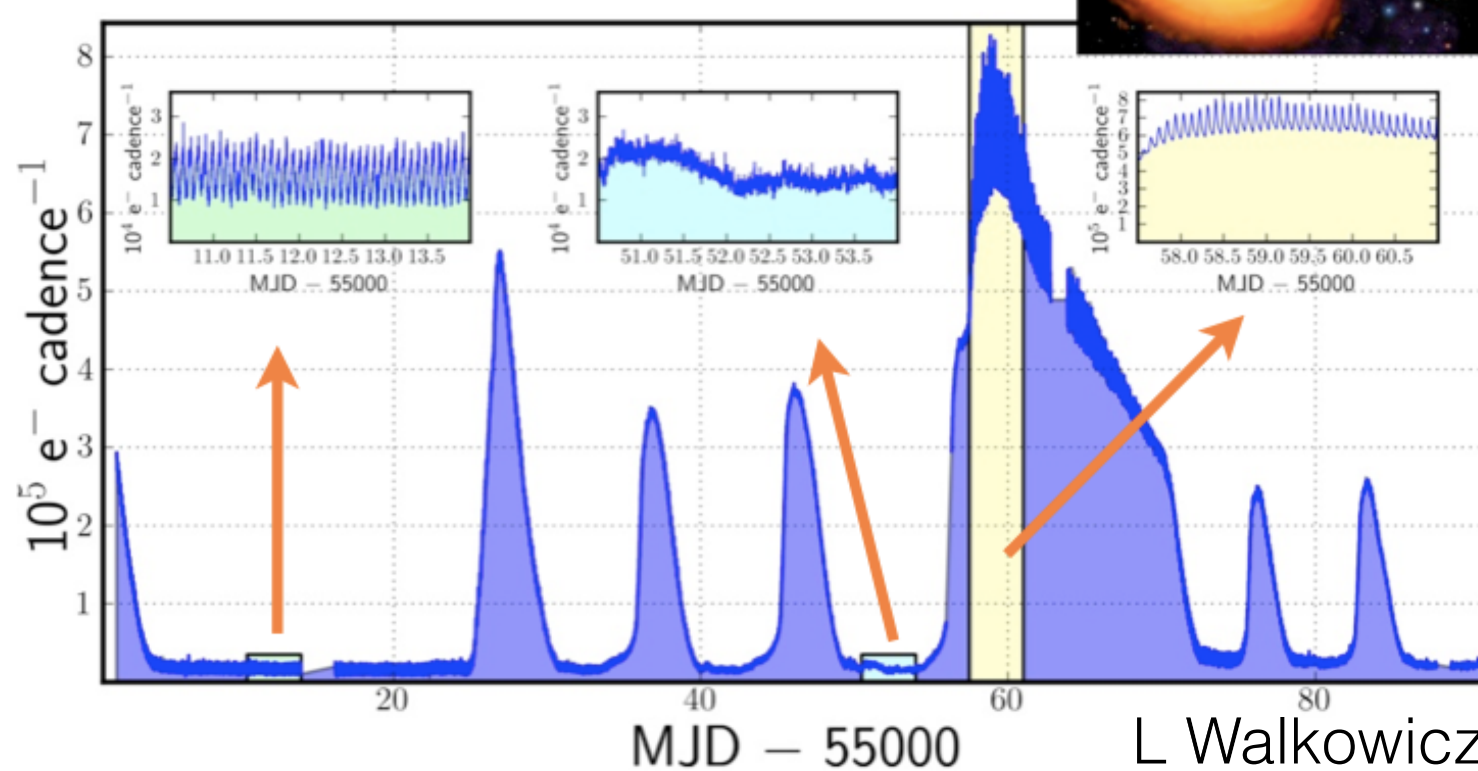
Time (3000+ days)



~100 days

Dwarf nova in the Kepler field





L Walkowicz

Time | Variable | Error

mjd | mag | magerr

Kepler - small area non-sparse

modified JD

JD = days since
12 noon 1 Jan -4712

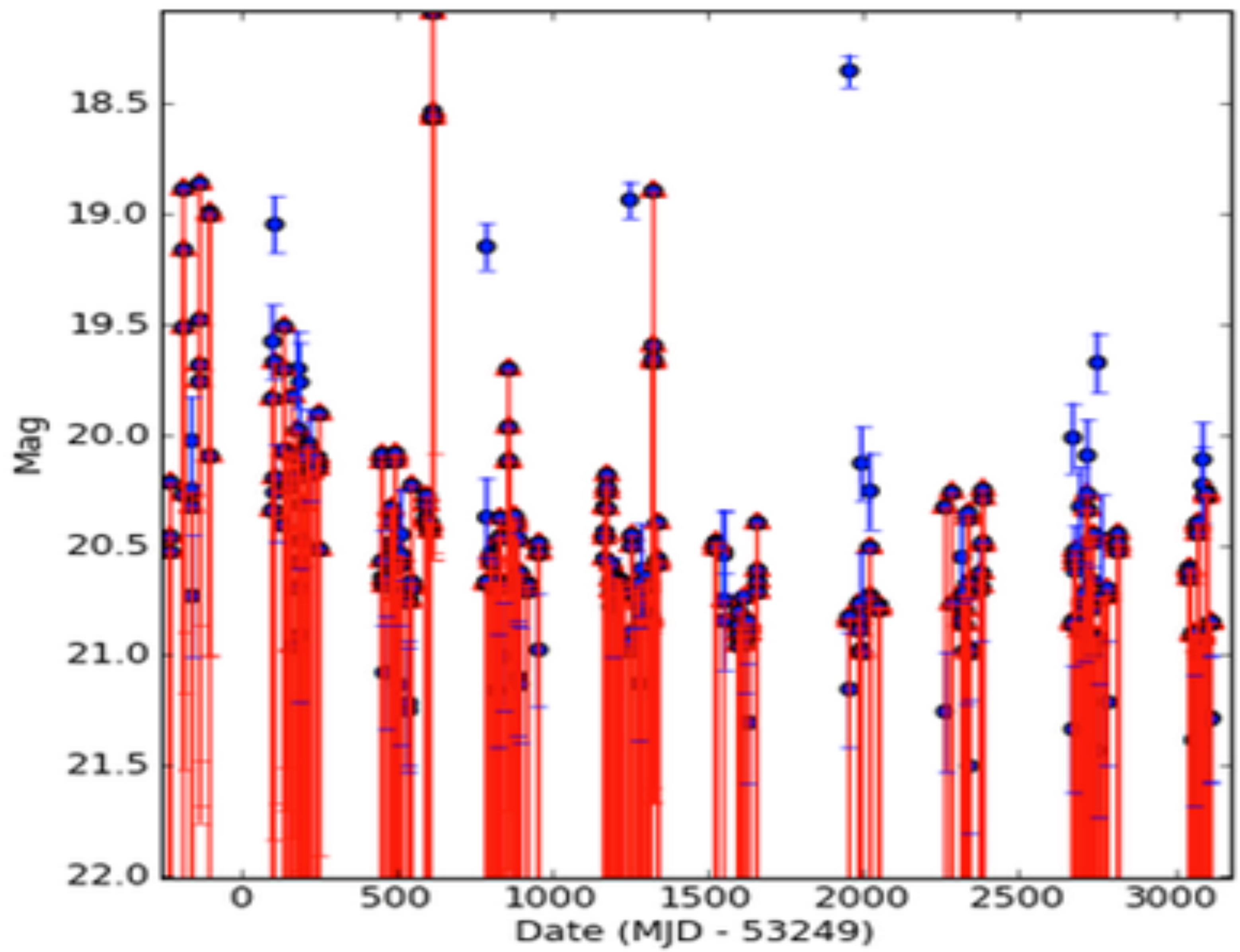
MJD = JD - 2400000.5

Typical time-series in astronomy

- DPOSS - large area, serendipitous overlap
- Kepler - small area non-sparse
- **CRTS - open filter, lumpy cadence for asteroids**
- **PTF/ZTF/Pan-STARRS/Gaia/LSST: multi filter, mixed**
- SKA/Radio
- Pulsars (timing arrays)

Properties of light-curves

- Gappy
- Irregular
- Heteroskedastic



Reasons:

- **expense, rotation/revolution of Earth, moon**
- **science objectives, weather, moon**
- **weather, moon, airmass**

errors ignored by many methods

What can we do with light-curves?

dreamstime.com

- Abstract them through generic statistical measures
- Use domain knowledge to look for characteristics
- See if they are periodic

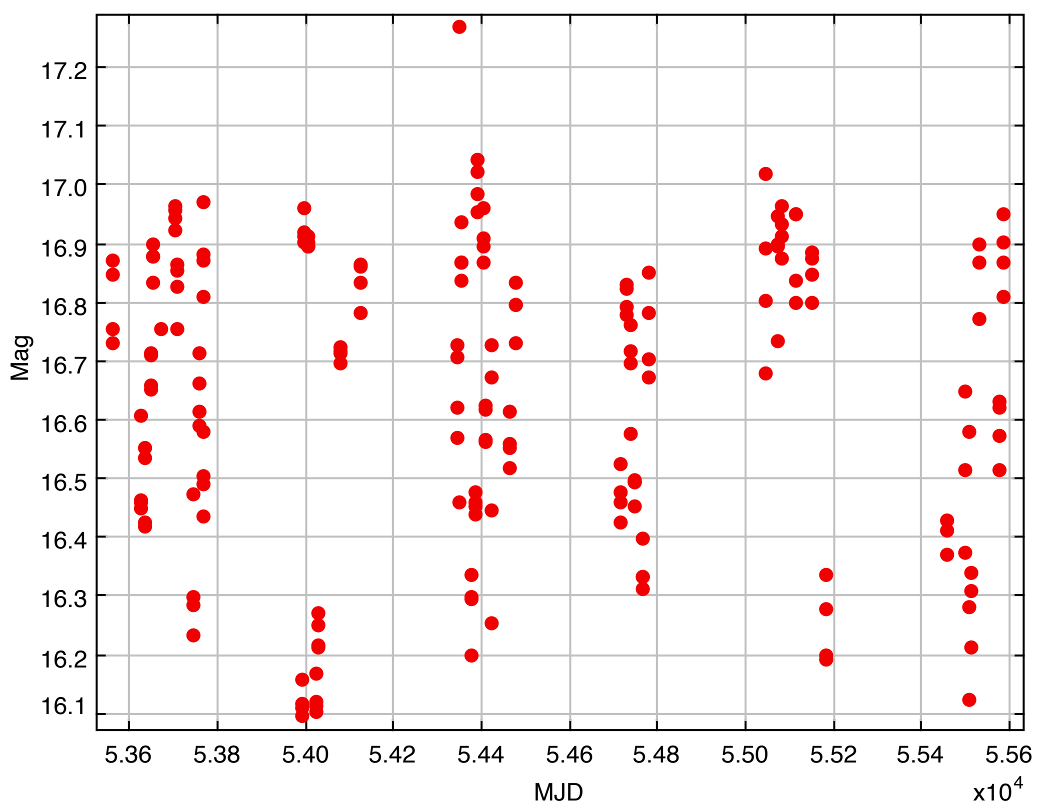
What can we do with light-curves?

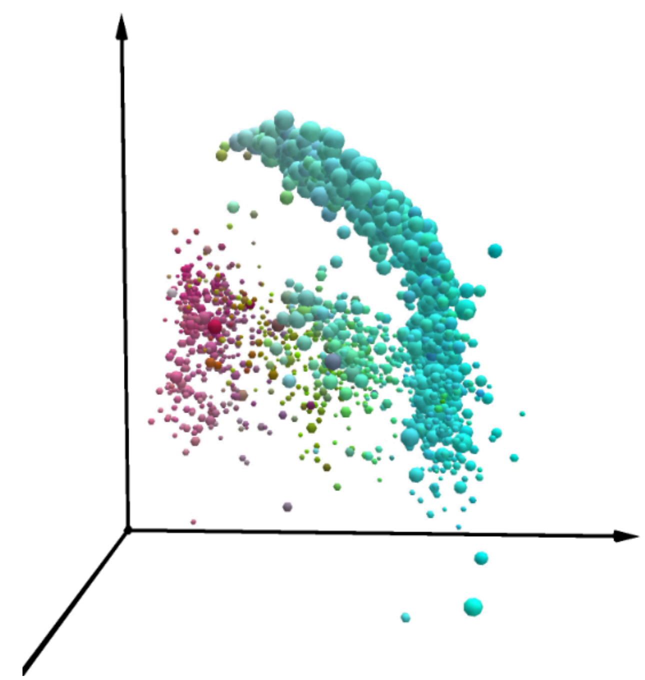
- Abstract them through generic statistical measures
- Use domain knowledge to look for characteristics
- See if they are periodic

Similarity measures:
dtwclust

Statistical features

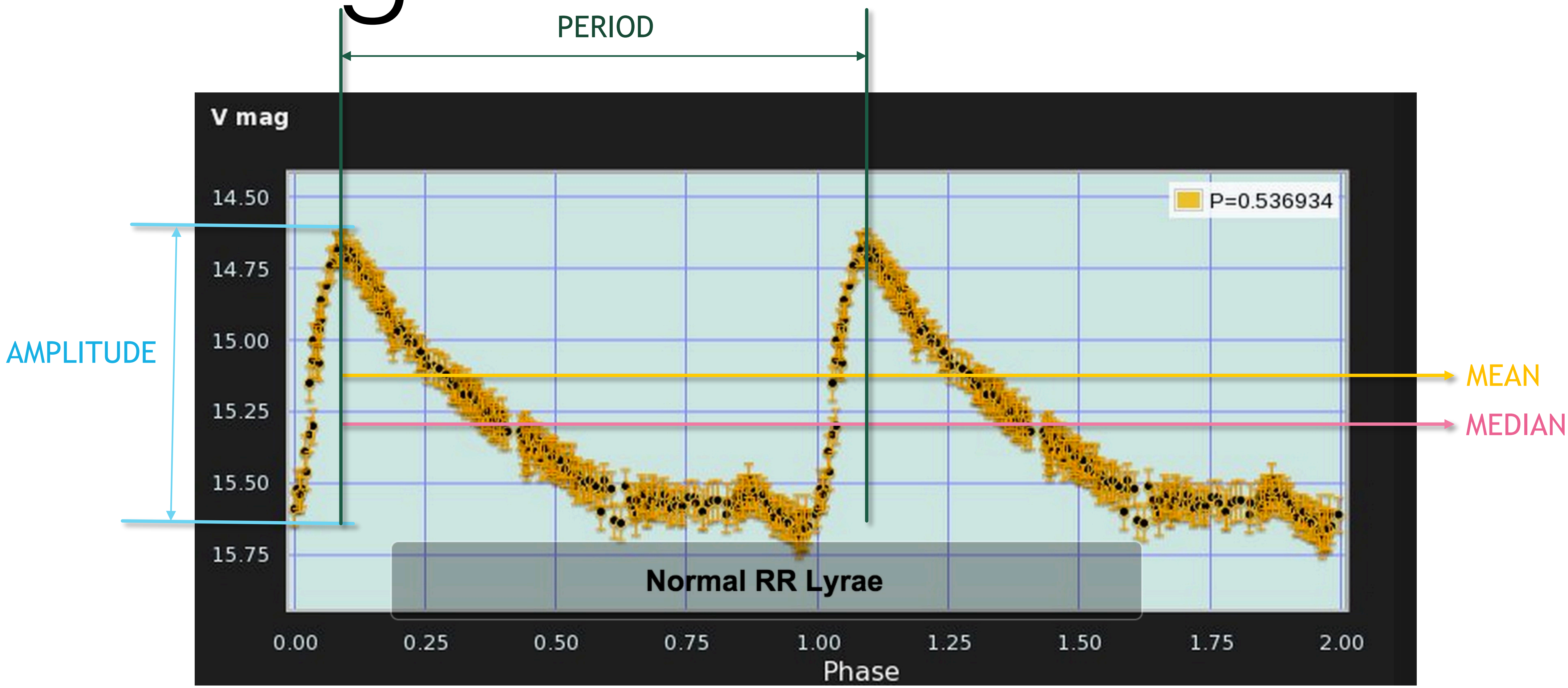
Compute features (statistical measures) for each light curve:
amplitudes, moments, periodicity, etc.
Converts heterogeneous light curves into homogeneous
feature vectors in the parameter space
Apply a variety of automated classification methods





Ashish Mahabal

Light-curve features



Statistical characteristics

Richards et al. (non-sparse OGLE-Hipparcos time-series) 2011

skew

small_kurtosis

std

beyond1std

stetson_j

stetson_k

max_slope

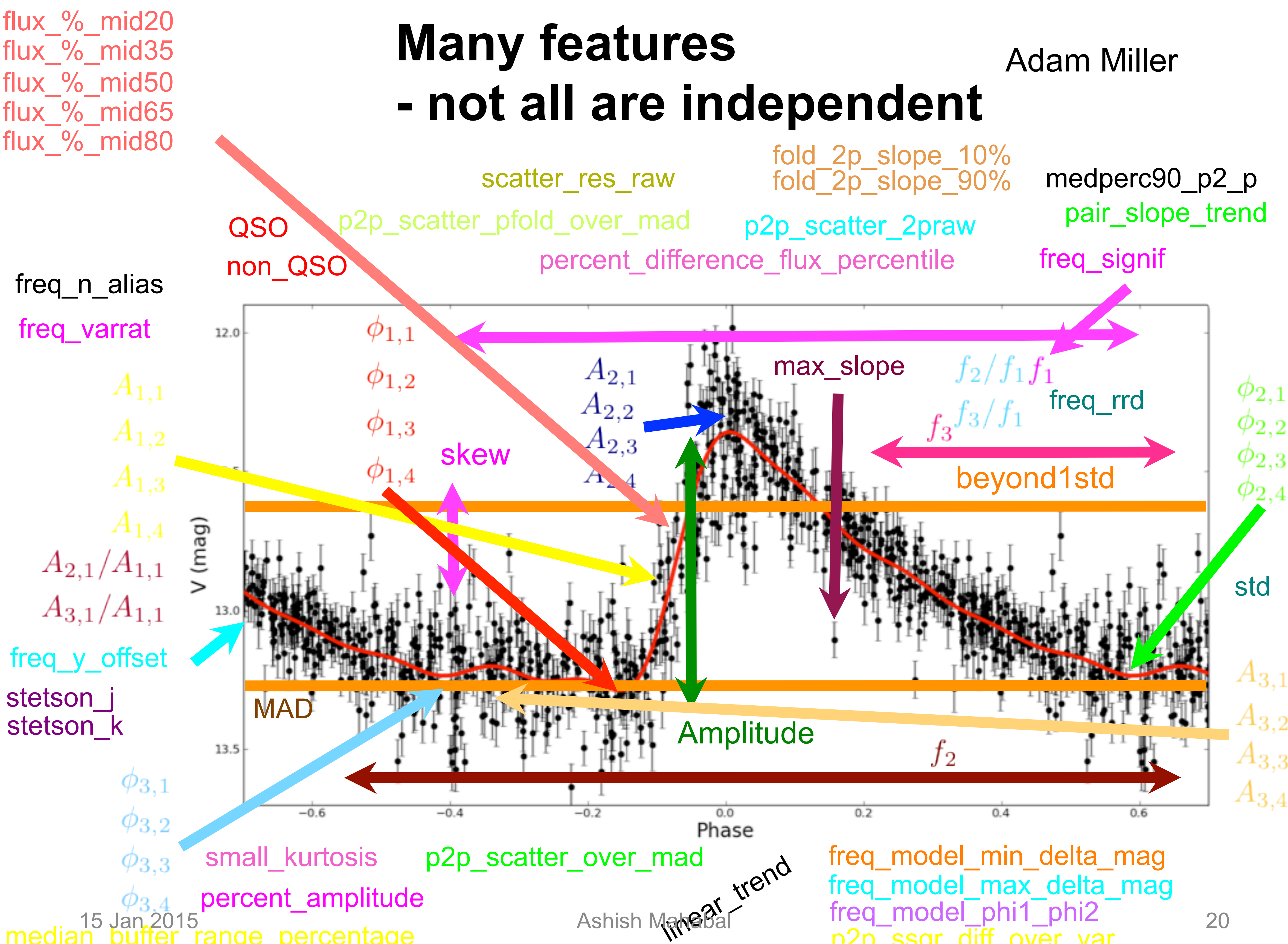
amplitude

Short name	Data type	Summary
amplitude	float	$0.5 * (mag_{max} - mag_{min})$
beyond1std	float	$p(\|(mag - \langle mag \rangle)\| > \sigma)$
flux_percentile_ratio_mid20	float	$(flux_{60} - flux_{40}) / (flux_{95} - flux_{5})$
flux_percentile_ratio_mid35	float	$(flux_{67.5} - flux_{32.5}) / (flux_{95} - flux_{5})$
flux_percentile_ratio_mid50	float	$(flux_{75} - flux_{25}) / (flux_{95} - flux_{5})$
flux_percentile_ratio_mid65	float	$(flux_{82.5} - flux_{17.5}) / (flux_{95} - flux_{5})$
flux_percentile_ratio_mid80	float	$(flux_{90} - flux_{10}) / (flux_{95} - flux_{5})$
linear_trend	float	b where mag = a * t + b
max_slope	float	$max(\|(mag_{i+1}-mag_{i})/(t_{i+1}-t_{i})\|)$
mad	float	$med(flux - flux_{med})$
median_buffer_range_percentage	float	$p(\|flux - flux_{med}\| < 0.1 * flux_{med})$
pair_slope_trend	float	$p(flux_{i+1} - flux_{i} > 0; i = n-30, n)$
percent_amplitude	float	$max(\|f_{max} - f_{med}\|, \|f_{min} - f_{med}\|)$
pdfp	float	$(flux_{95} - flux_{5}) / flux_{med}$
qso	4x1	var_{qso}
skew	float	μ_3/σ^3
small_kurtosis	float	μ_4/σ^4
std	float	σ
stetson_j	float	$var_j(mag)$
stetjon_k	float	$var_k(mag)$

lightcurve characterization service: http://nirgun.caltech.edu:8000/

Many features
- not all are independent

Adam Miller



15 Jan 2015 Ashish Mahabal 20

Stetson Stats

Welch-Statson
1996PASP..108..851S

$$I=\sqrt{\frac{1}{n(n-1)}}\sum_{i=1}^{n}\left(\frac{b_i-\bar{b}}{\sigma_{b,i}}\right)\left(\frac{v_i-\bar{v}}{\sigma_{v,i}}\right),$$

Pairwise observations in 2 filters

$$J=\frac{\Sigma_{k=1}^{n} w_k \operatorname{sgn}(P_k)\sqrt{|P_k|}}{\Sigma_{k=1}^{n} w_k},$$

Pairwise observations (single filter)

$$K=\frac{1/N \ \ \Sigma_{i=1}^{N}|\delta_i|}{\sqrt{1/N \ \ \Sigma_{i=1}^{N}\delta_i^2}},$$

No pairing required

$$L=\left(\frac{JK}{0.798}\right)\left(\frac{\Sigma w}{w_{\rm all}}\right).$$

Combined for thresholding

CRTS variables

- 150M sources from a few thousand “fields”
- ~5.5M variables after filtering using per field J
- ~50K periodic (LS False Alarm Probability < 10^-5; M_t thresholds)

Drake et al. 2014

  - 15 classes

M_t: Fraction of time below median (Kinemuchi et al. 2006)

LS by Many names

$$\phi(t) = A \sin \omega t + B \cos \omega t + C.$$

sines + cosines < n
generalized version fits for mean
(rather than using mean = 0 through subtraction)

- Lomb-Scargle periodogram is a least squares sinusoid fit (Least Squares Spectral Analysis)

- Matching Persuit

Entropy based period finding

$$H_0 = -\sum_{i=1}^{k} \mu_i \ln(\mu_i) \quad \forall \mu_i \neq 0,$$

Counts in k-partitions after phasing

1-day aliasing!

$$H_c = \sum_{i,j} p(m_i, \phi_j) \ln\left(\frac{p(\phi_j)}{p(m_i, \phi_j)}\right),$$

Counts in partitions after
phasing in time
and binning in mags

Graham et al. 2013

Q: Amplitude variations

$$Q = \frac{(\mathrm{RMS}^2_{\mathrm{resid}} - \sigma^2)}{(\mathrm{RMS}^2_{\mathrm{raw}} - \sigma^2)}, \qquad (6)$$

where $\mathrm{RMS}_{\mathrm{raw}}$ and $\mathrm{RMS}_{\mathrm{resid}}$ are the RMS values of the raw light curve and the phase subtracted light curve, respectively, whereas σ is the estimated uncertainty including the systematics (e.g., Section 3.3). Testing on si-

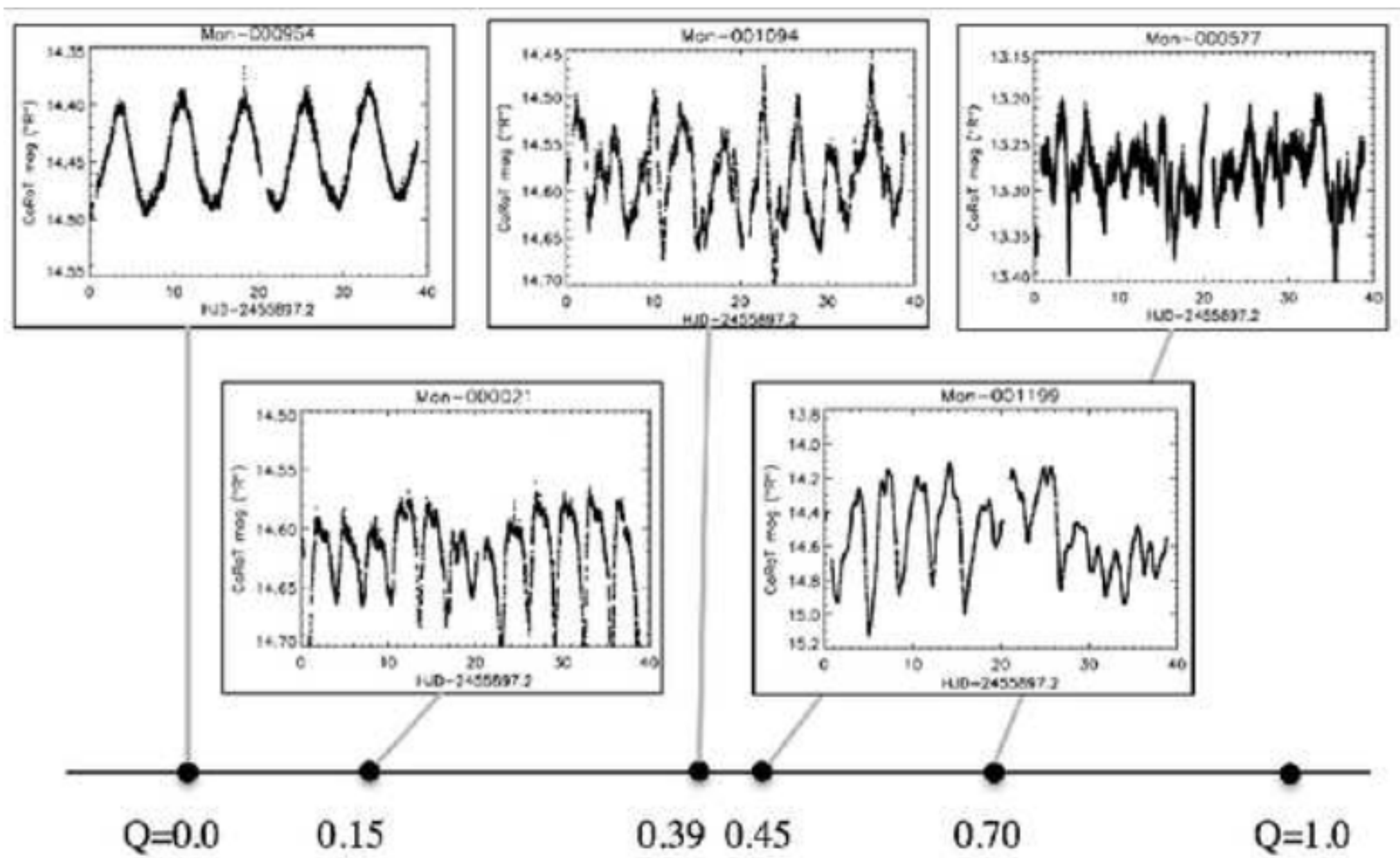


FIG. 29.— *CoRoT* light curves with representative values of the Q parameter, ranging from periodic (Q=0–0.15) to quasi-periodic (Q=0.15–0.5), to aperiodic $Q > 0.5$.

Cody et al. 2014

M: Bursters and dippers

$$M = (< d_{10\%} > - d_{\mathrm{med}})/\sigma_d, \quad (7)$$

where $< d_{10\%} >$ is the mean of all data at the top and bottom decile of light curve, d_{med} is the median of the entire light curve, and σ_d is its overall RMS.

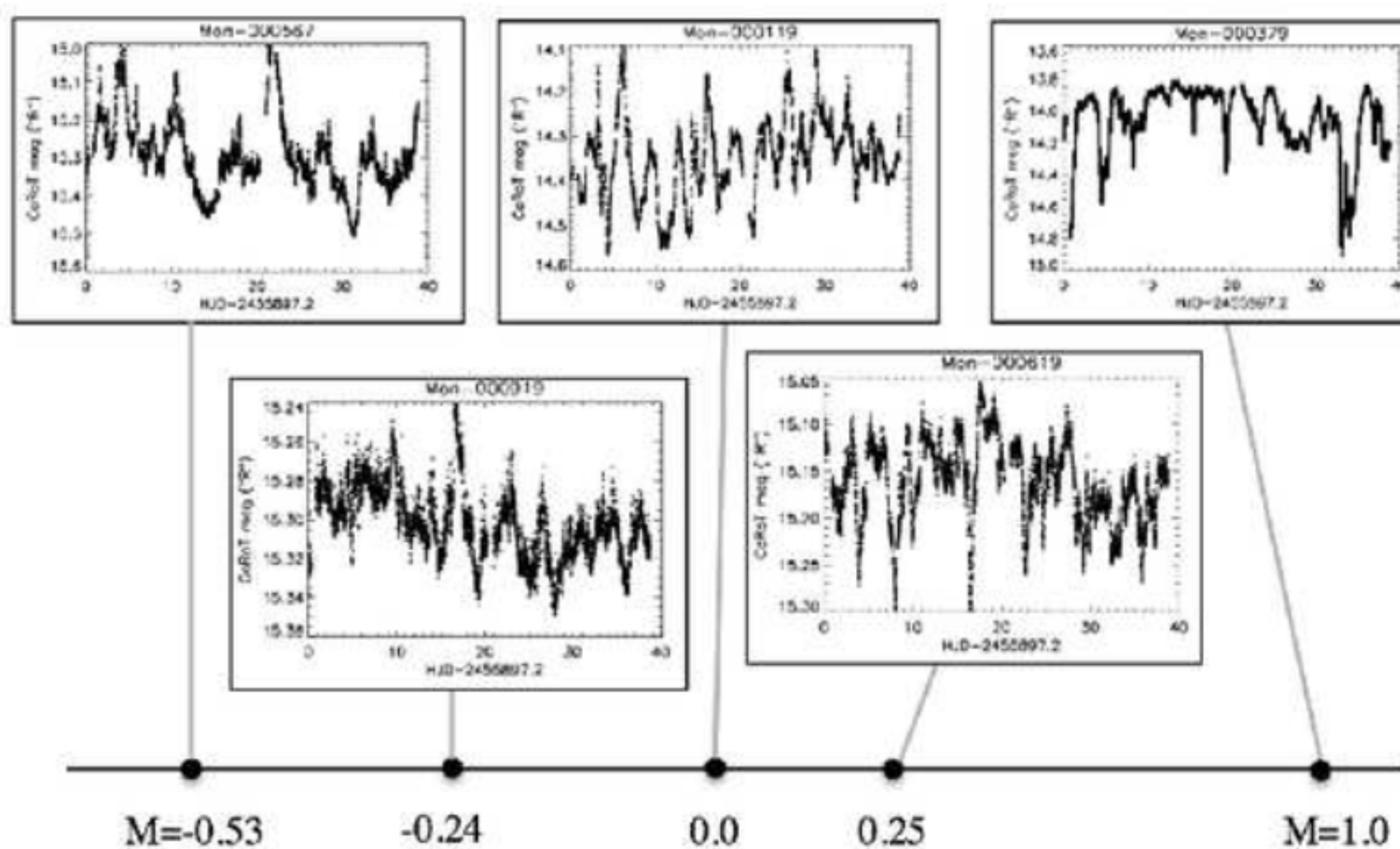


FIG. 30.— *CoRoT* light curves with representative values of the M parameter, ranging from bursting ($M < -0.25$) to symmetric (M=-0.25–0.25), to dipping $M > 0.25$.

Cody et al. 2014

Q-M plane

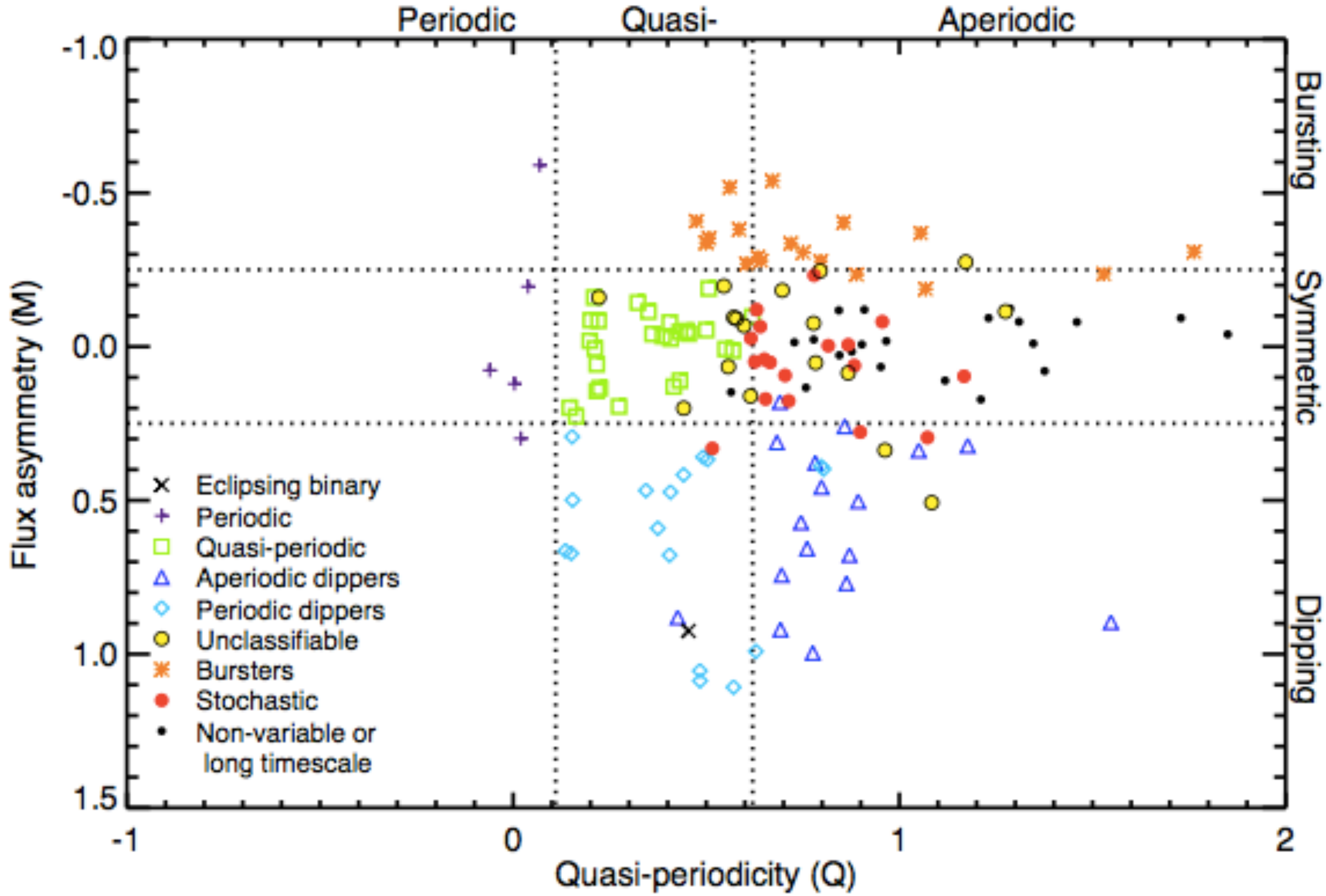


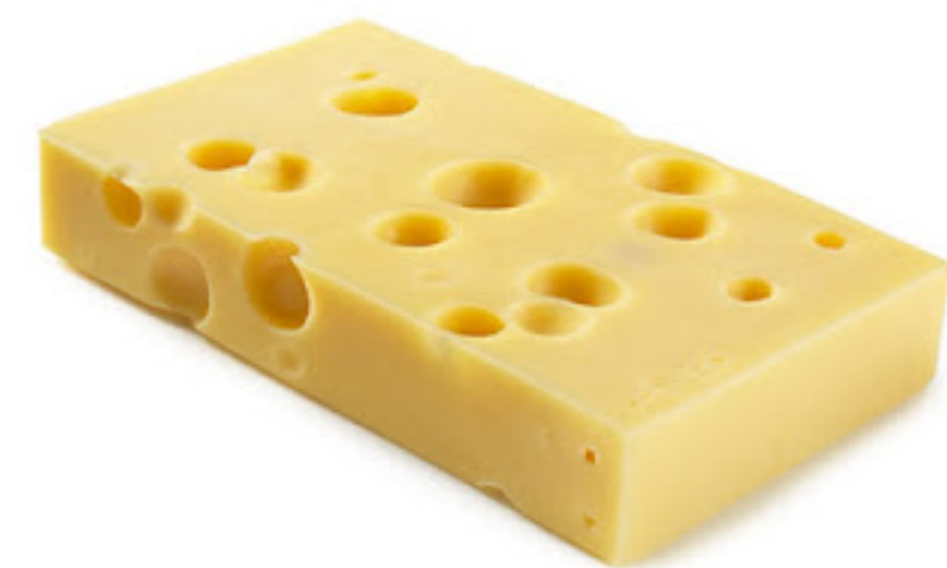
FIG. 31.— *Top*: Light curve morphology classes, as divided by the quasi-periodicity (Q) and flux asymmetry (M) parameters for optical light curves from *CoRoT* in our disk-bearing sample. Color coding indicates the variability classification chosen by eye, before statistical assessment. The eclipsing binary is not strictly periodic because its light curve contains aperiodic fluctuations out of eclipse. *Bottom*: Same

Cody et al. 2014

Challenge: A Variety of Parameters

- Discovery: magnitudes, delta-magnitudes
- Contextual:
  - Distance to nearest star
  - Magnitude of the star
  - Color of that star
  - Normalized distance to nearest galaxy
  - Distance to nearest radio source
  - Flux of nearest radio source
  - Galactic latitude
- Follow-up
  - Colors (g-r, r-I, i-z etc.)
- Prior classifications (event type)
- **Characteristics from light-curve**
  - **Amplitude**
  - **Median buffer range percentage**
  - **Standard deviation**
  - **Stetson k**
  - **Flux percentile ratio mid80**
  - **Prior outburst statistic**

Not all parameters are always present leading to swiss-cheese like data



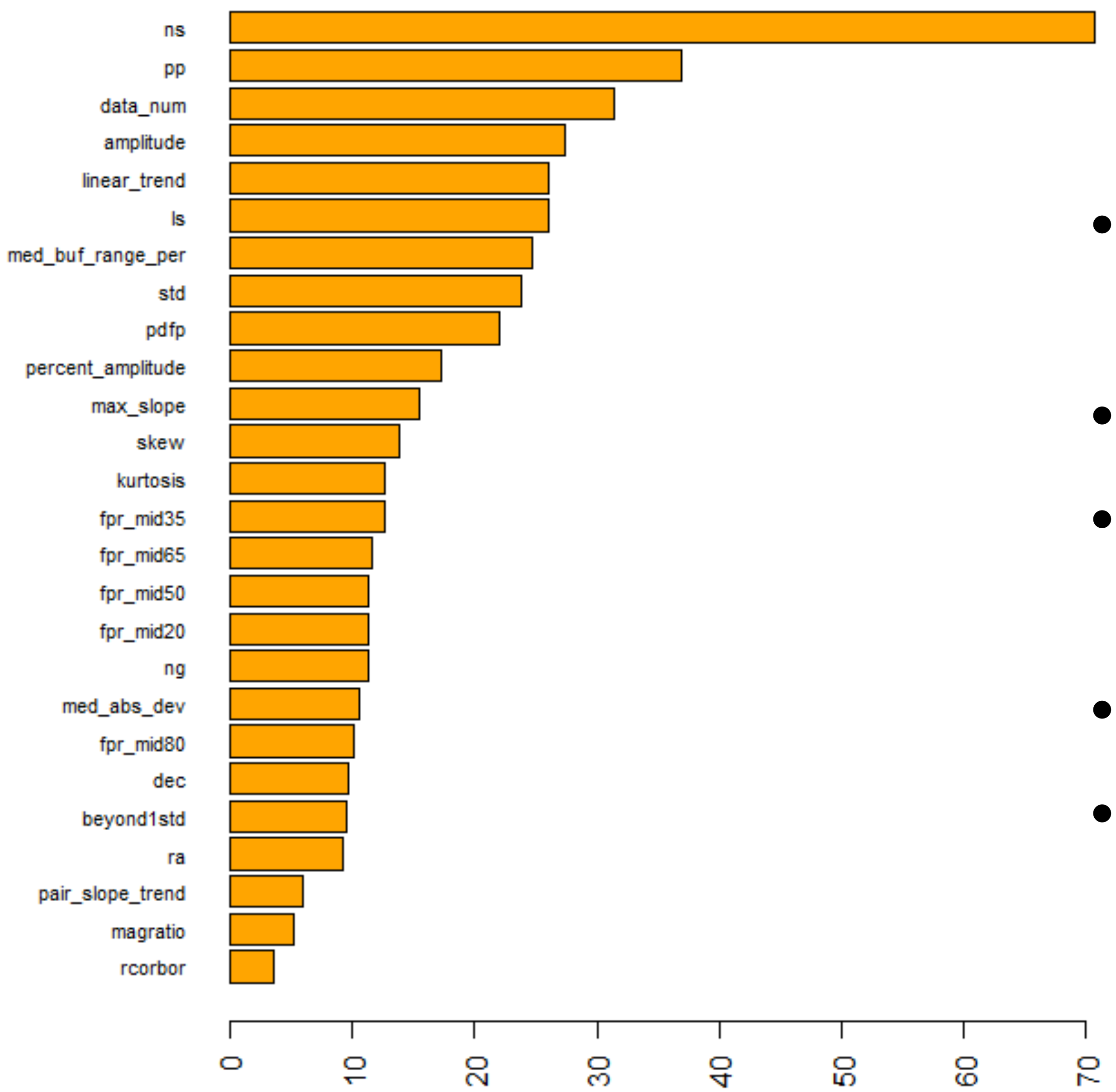
http://ki-media.blogspot.com/

Measures from Feigelson and Babu (Graham)

New lightcurve-based parameters: (Faraway)

•Whole curve measures
•Fitted curve measures
•Residual from fit measures
•Cluster measures
•Other

Feature selection strategy

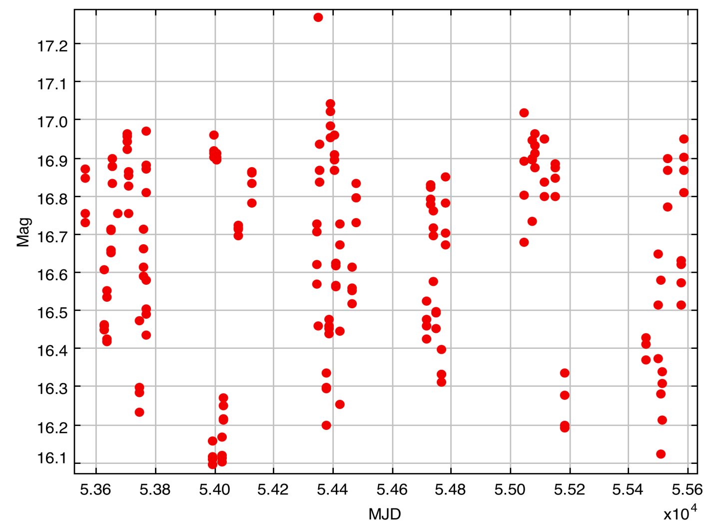


- Fast Relief Algorithm (wt and threshold)
- Fisher Discriminant Ratio
- Correlation based Feature Selection
- Fast Correlation Based Filter
- Multi Class Feature Selection

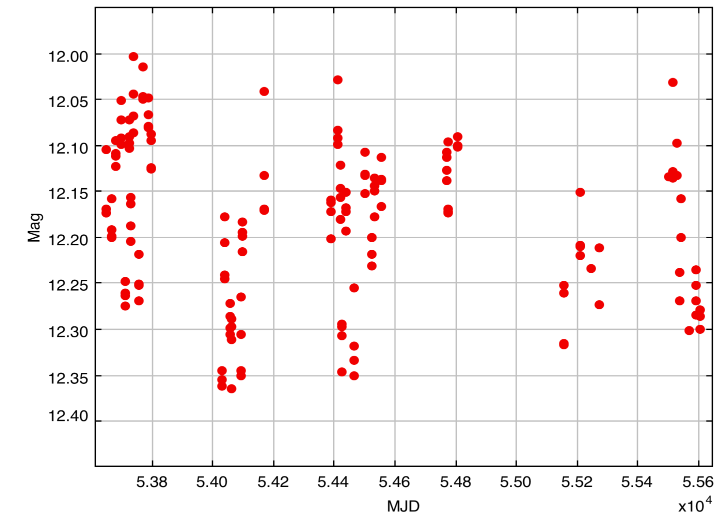
Donalek, .., Mahabal, … arxiv:1310.1976

Features for RR Lyrae and W UMa

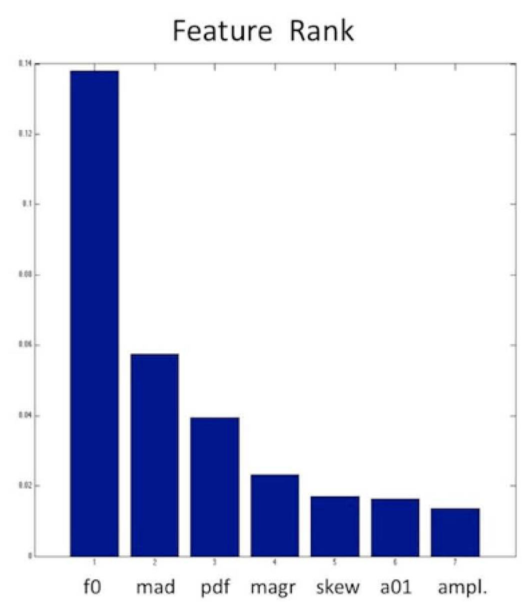
Rank features in the order of classification quality for a given classification problem, e.g., RR Lyrae vs. WUMa

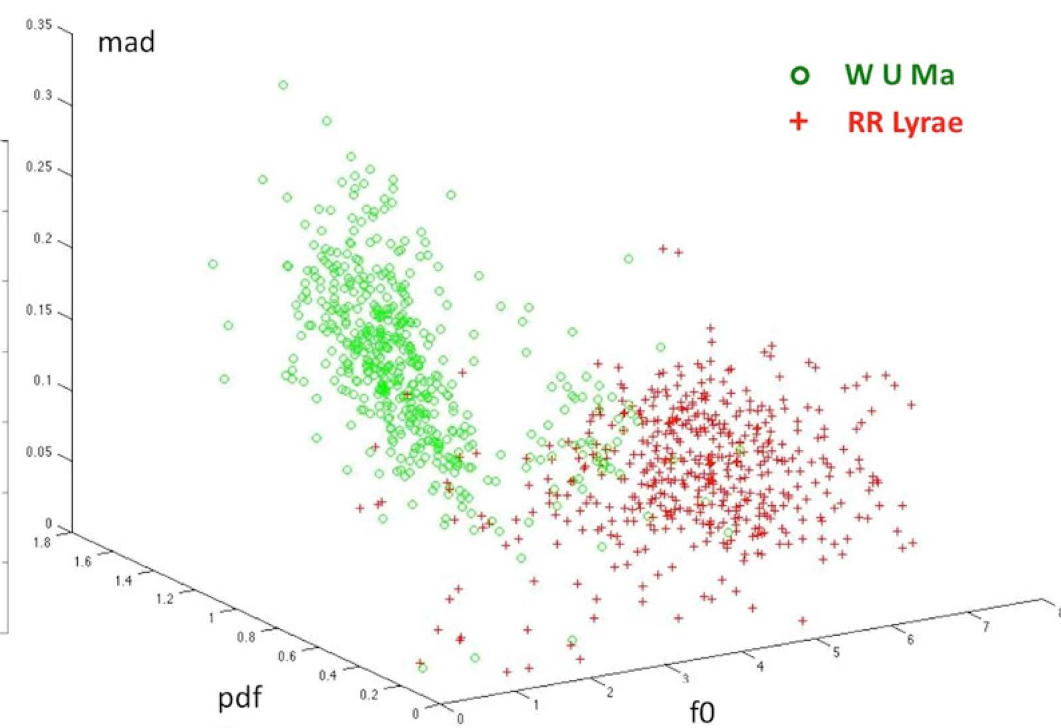


RR Lyrae



Eclipsing binary (W U Ma)





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

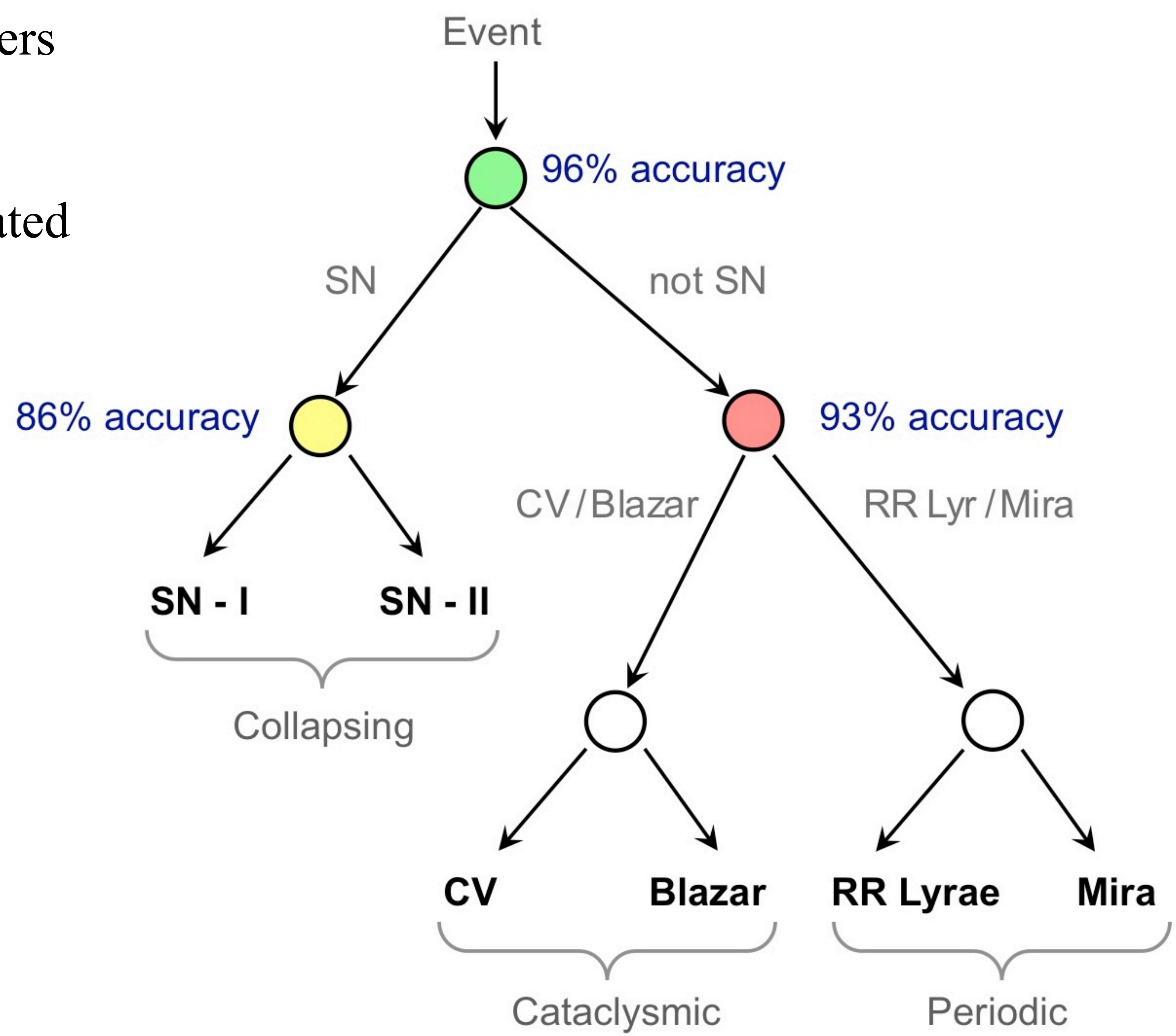
Fit the means within the groups (up to 4 measurements); and then take the
logged SD of the residuals from this fit; the max absolute residuals from this fit;
total variation of curve based on group means scaled by range of observation

A Hierarchical Approach to Classification

Different types of classifiers perform better for some event classes than for the others

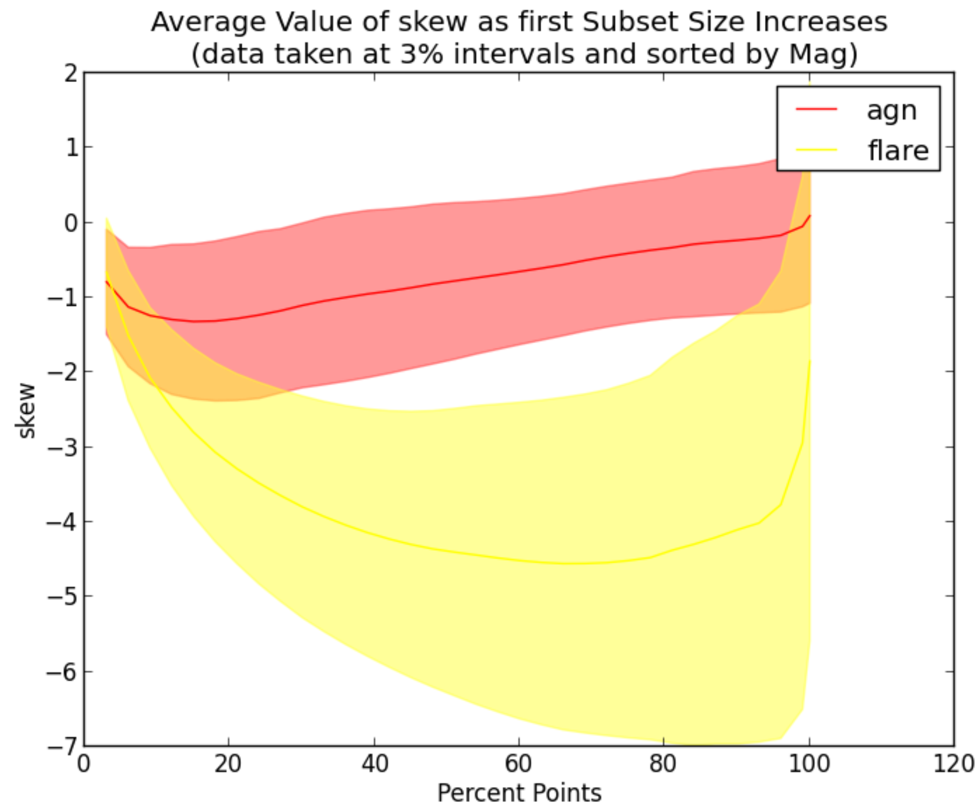
We use some astrophysically motivated major features to separate different groups of classes

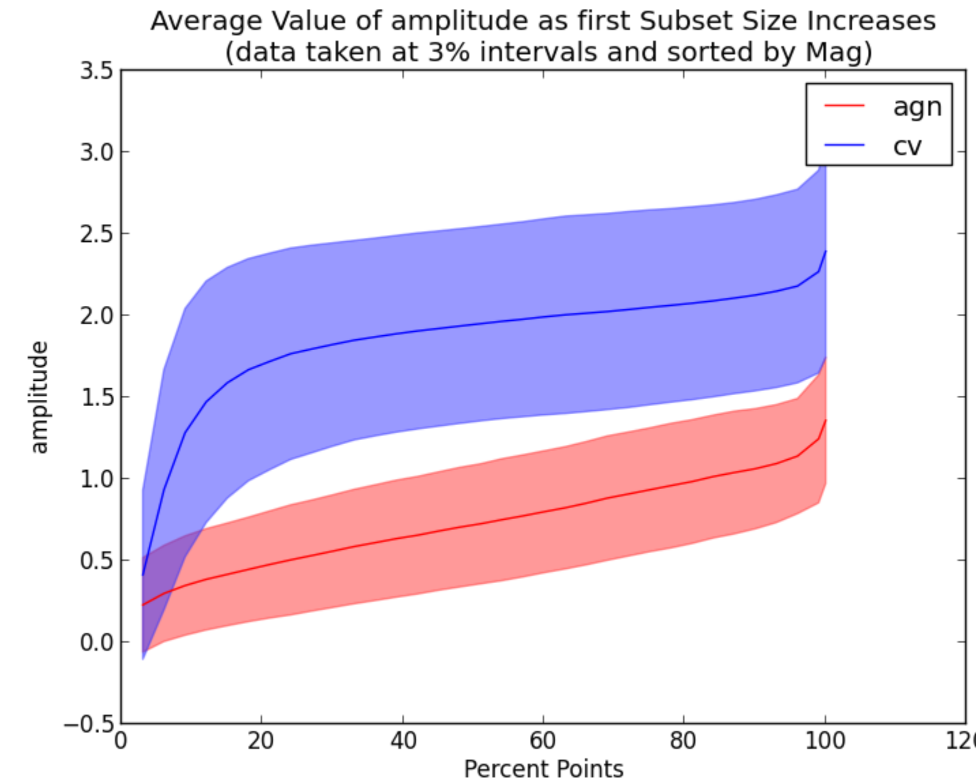
Proceeding down the classification hierarchy every node uses those classifiers that work best for that particular task

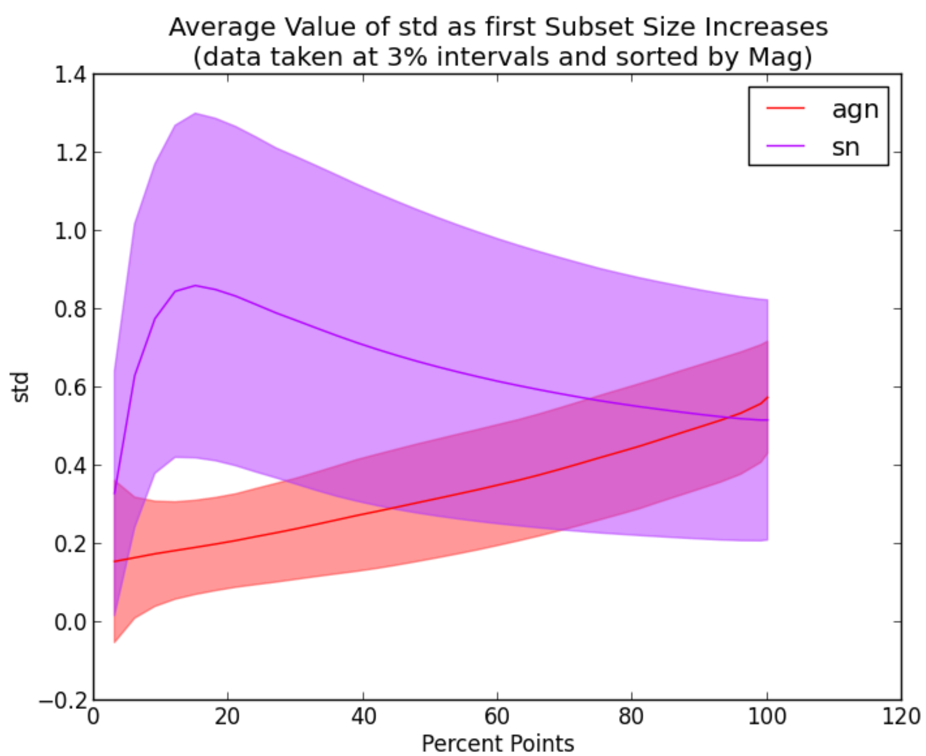


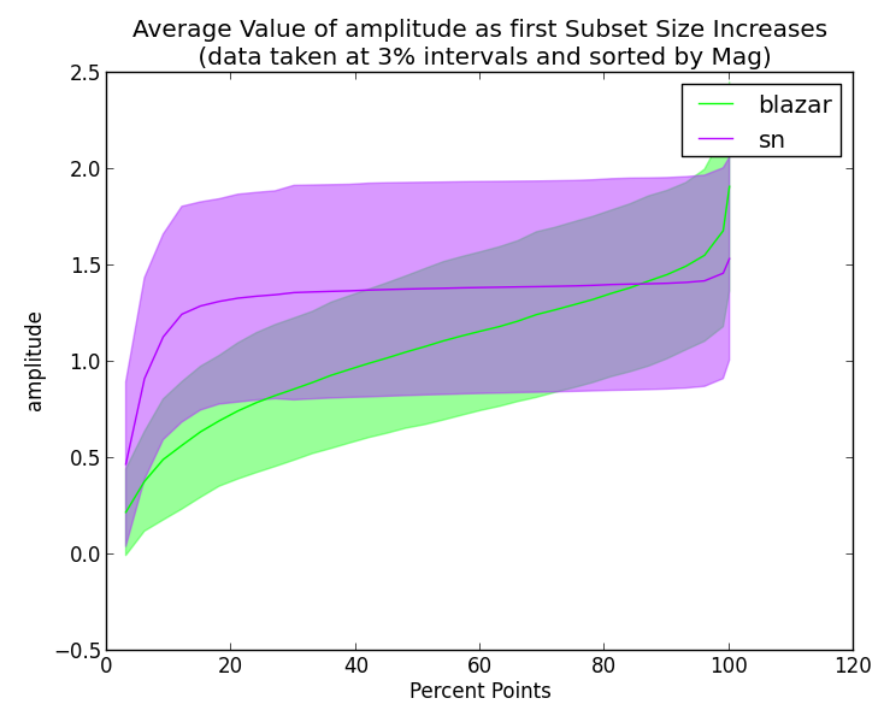
Using Discriminating Features for Brokering

Chengyi Lee





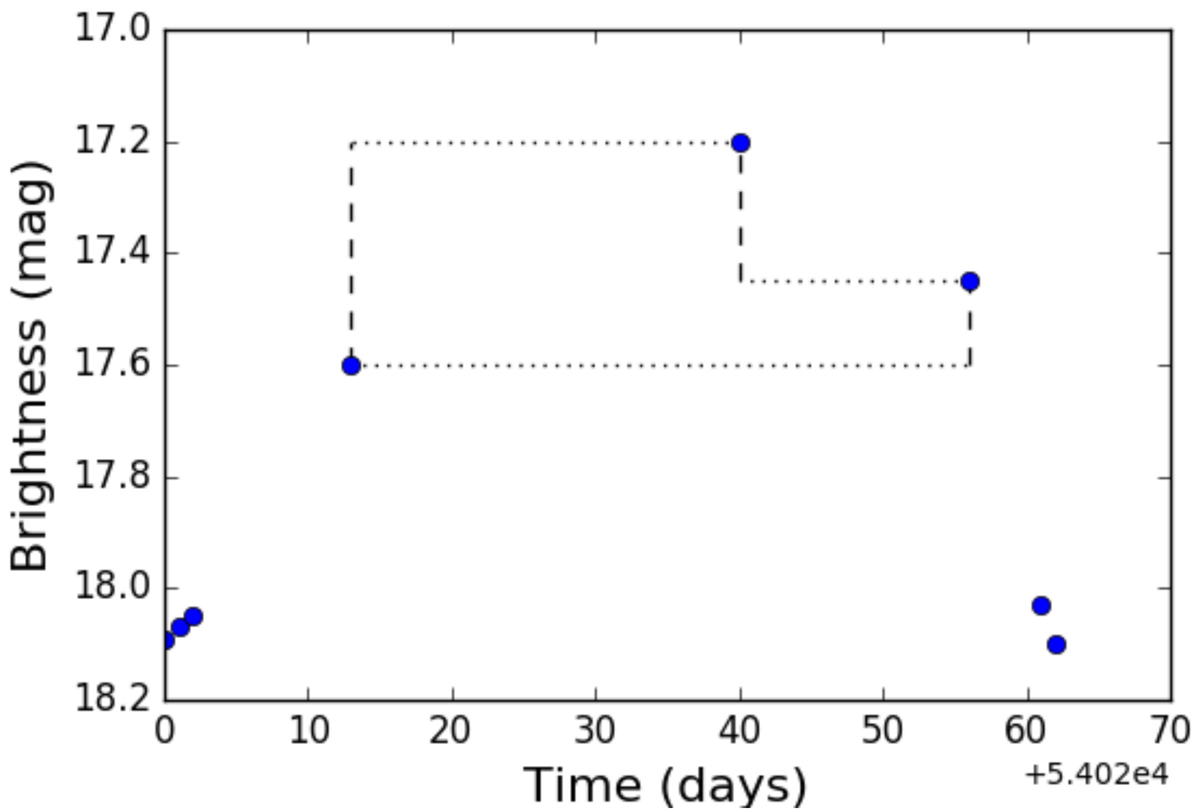




You can not step into the same river twice.

(dmdt) Image representation

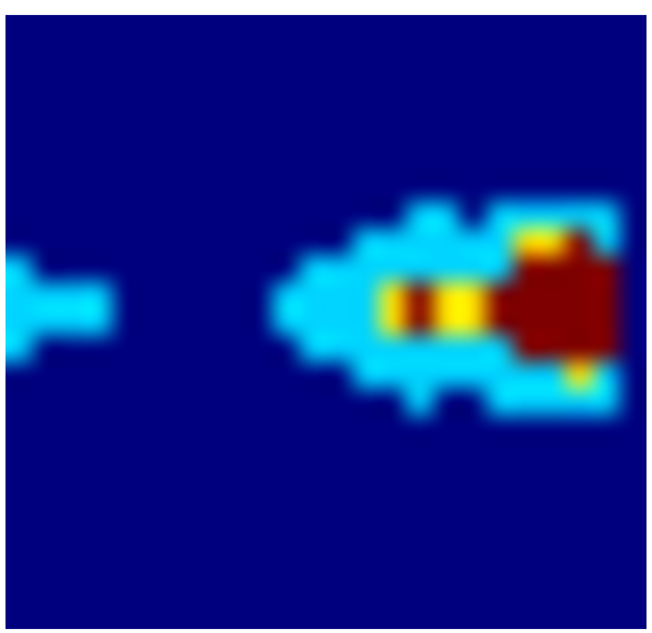
Mahabal et al., 2017

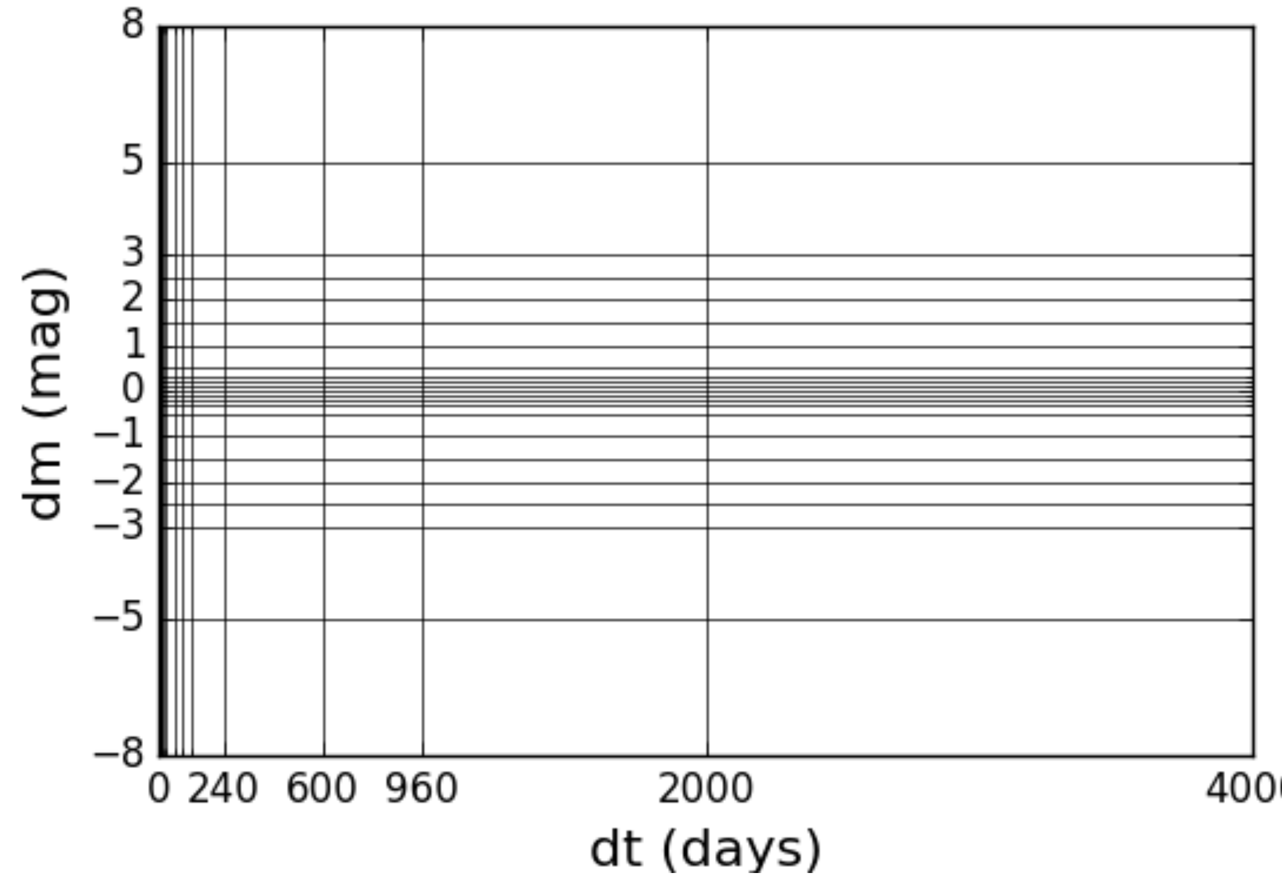


light curve with n points

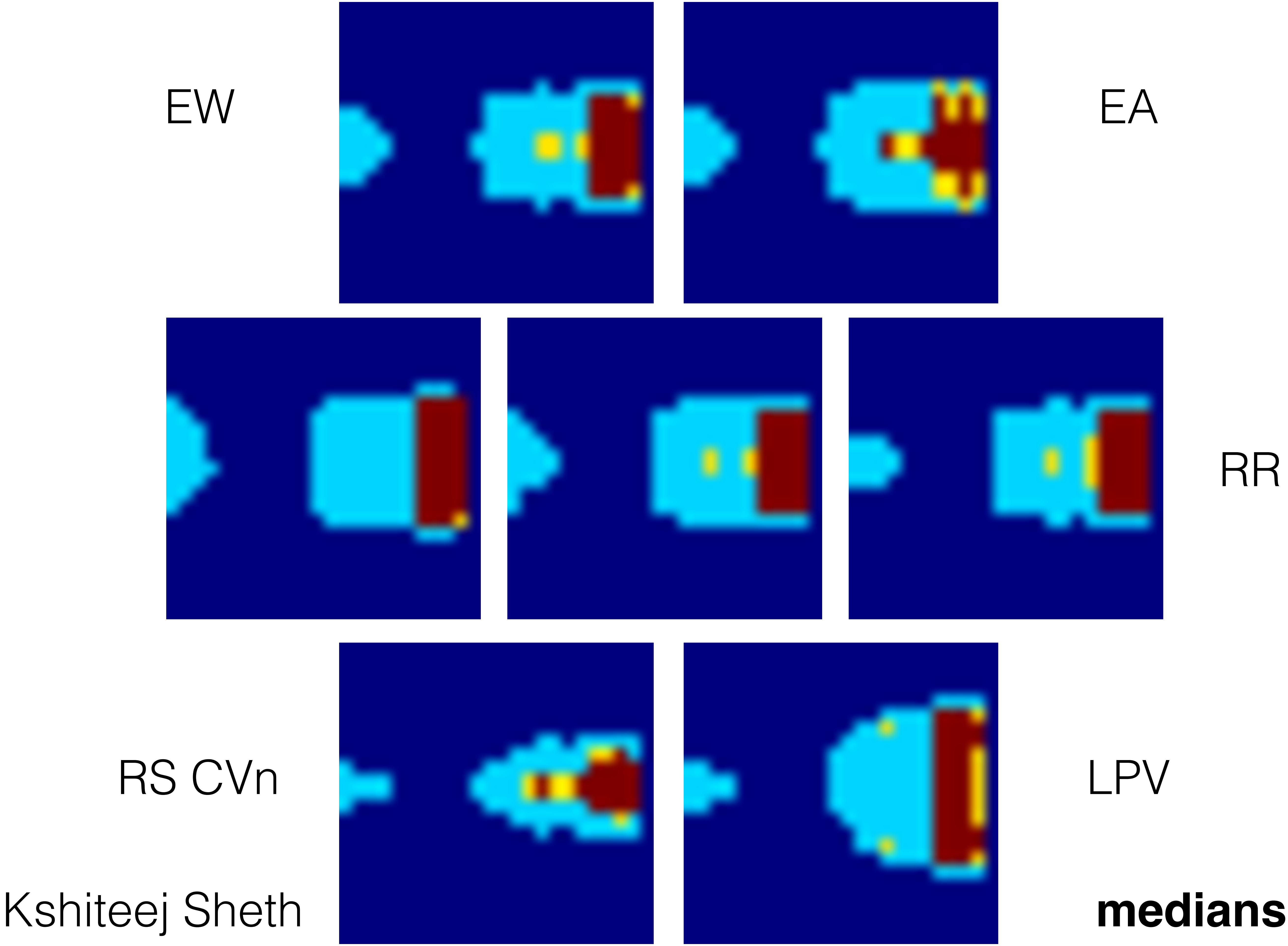
23 x 24 output grid

n * (n-1)/2 points



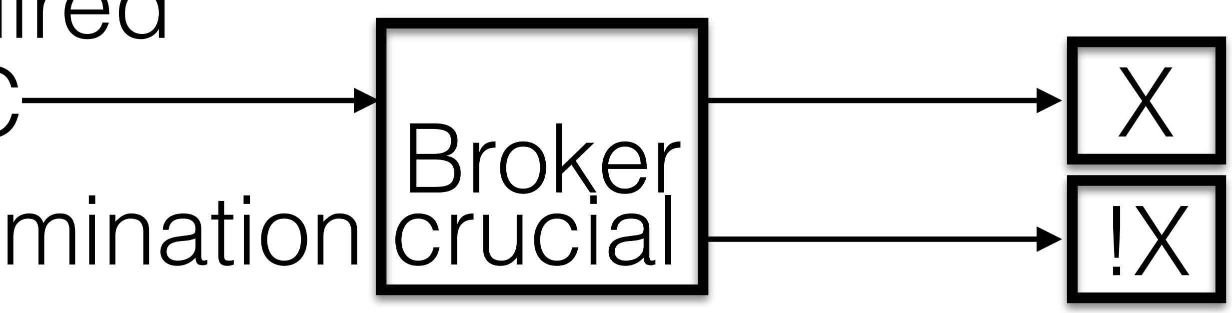


Area equalized pixels


EW
EA
RR
RS CVn
LPV
Kshiteej Sheth
medians

Binary Broker(s)

- Using features to tell classes apart - one class at a time
- Speed required
- Rarity determination crucial

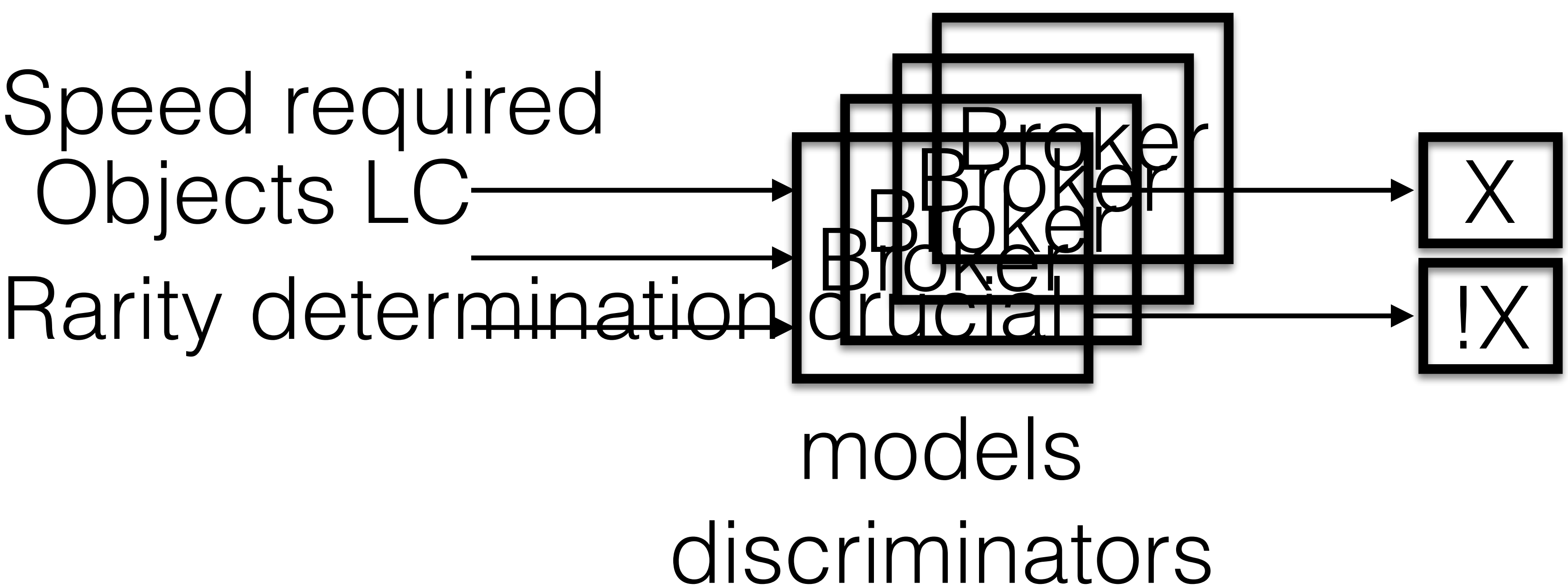


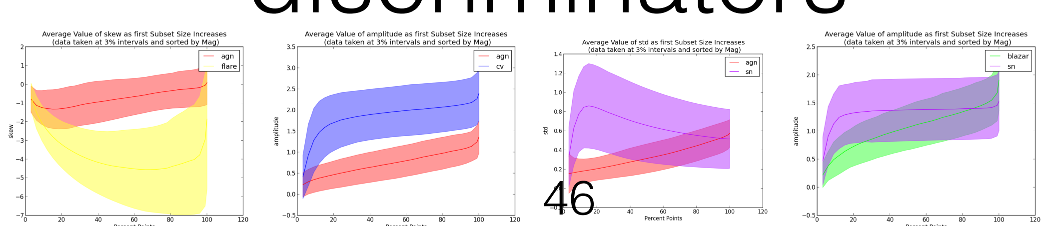
Binary Broker(s)

- Using features to tell classes apart - one class at a time

- Speed required

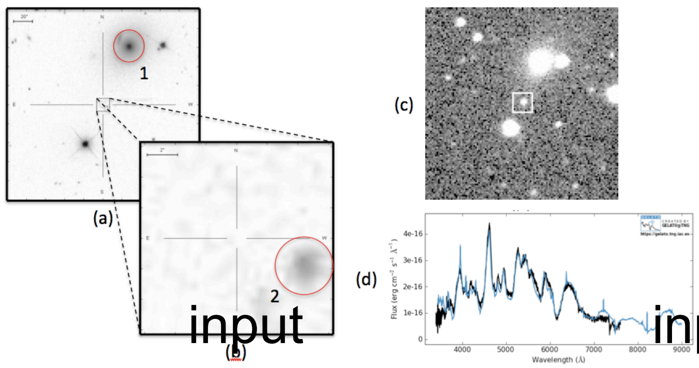
- Rarity determination crucial





Ashish Mahabal

Binary Brokers



Inputs:
Light-curves
Nearby objects
Archival catalogs

input
Not C1
C1

input
Not C2
C2

input
Not C3
C3

...

Modular

input
Not C1
C1

input
Not C2
C2

input
Not C3
C3

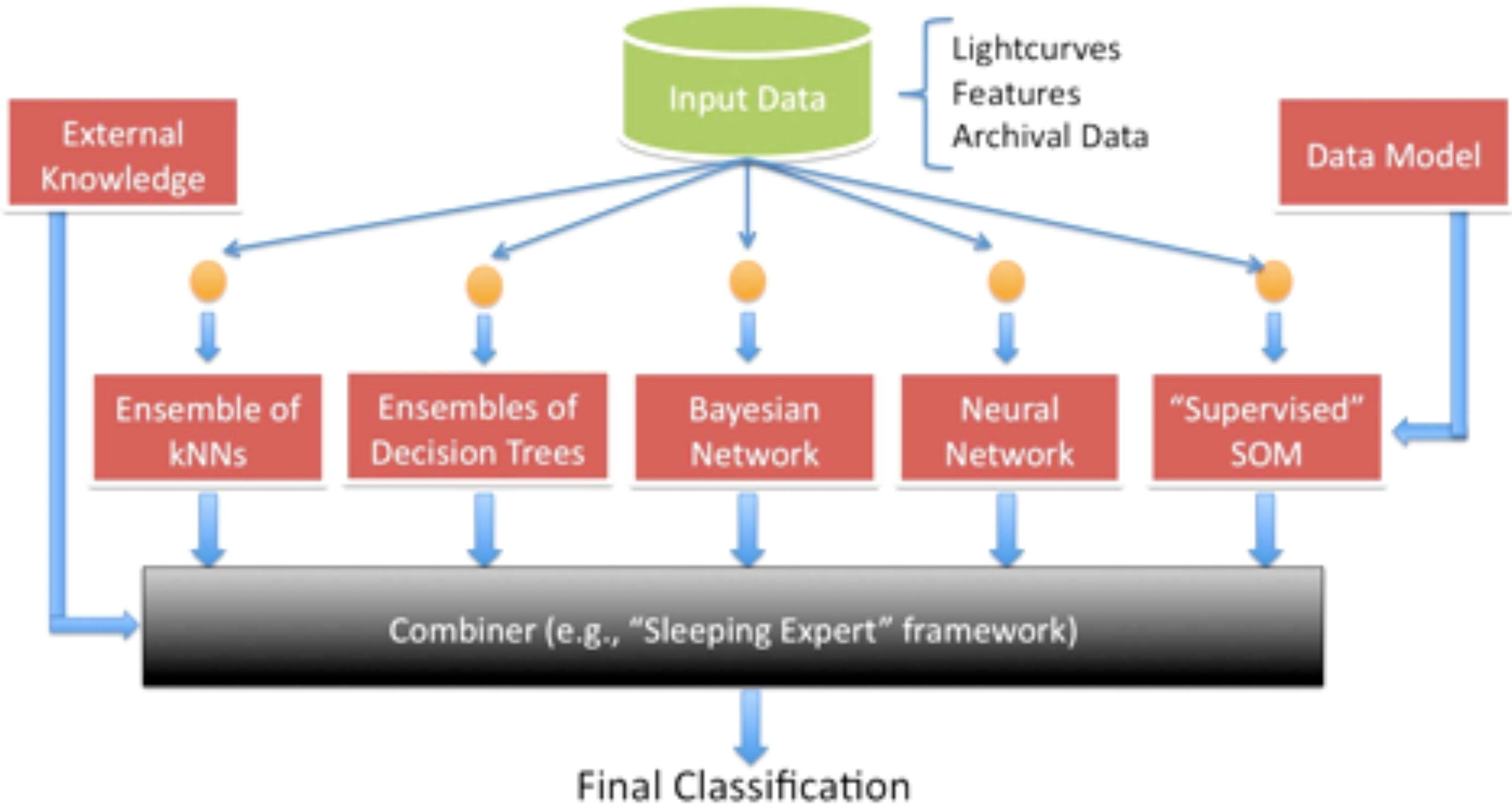
+

input
Not C4
C4

input
Not C5
C5

Extendible

Metaclassification: An optimal combining of classifiers



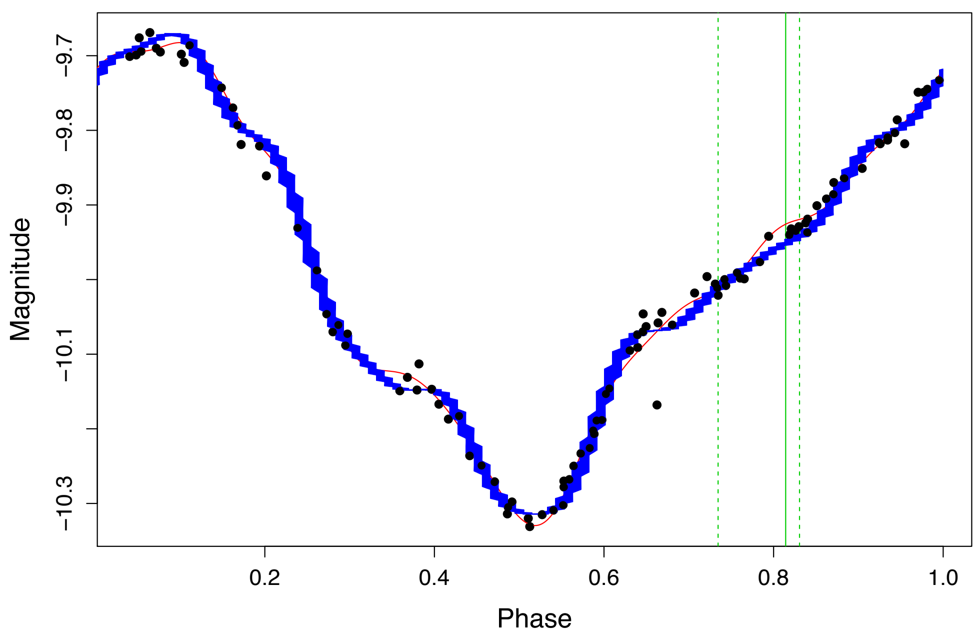
Exploring a variety of techniques for an optimal classification fusion:
Markov Logic Networks, Diffusion Maps, Multi-Arm Bandit,
Sleeping Expert…

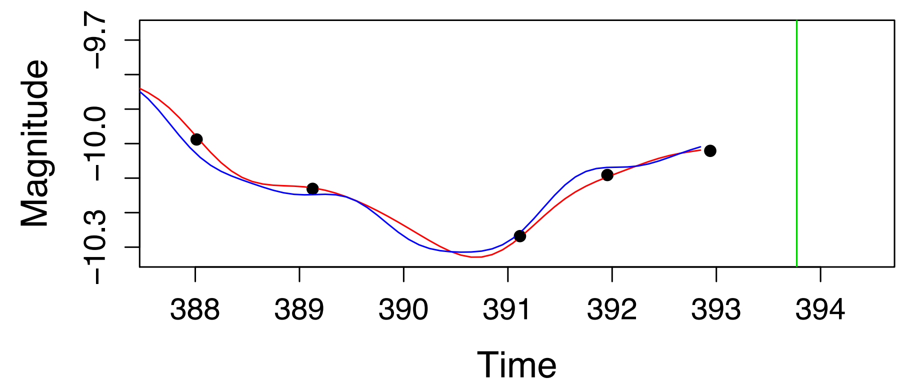
Mahabal, Donalek

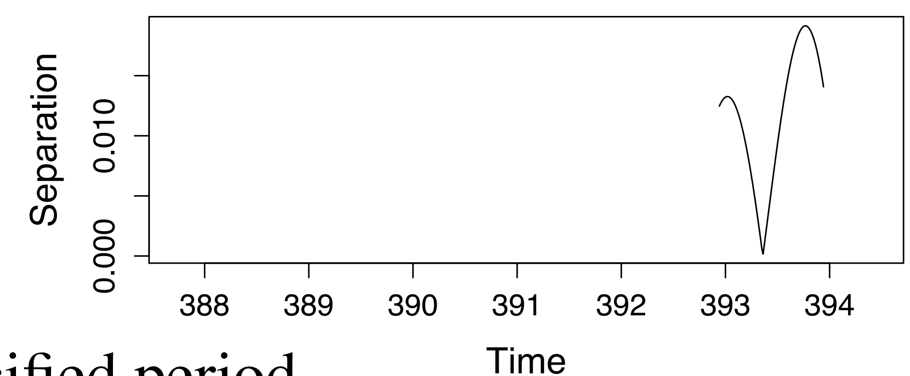
Scheduling observations

Toy Cepheid example

D Jones





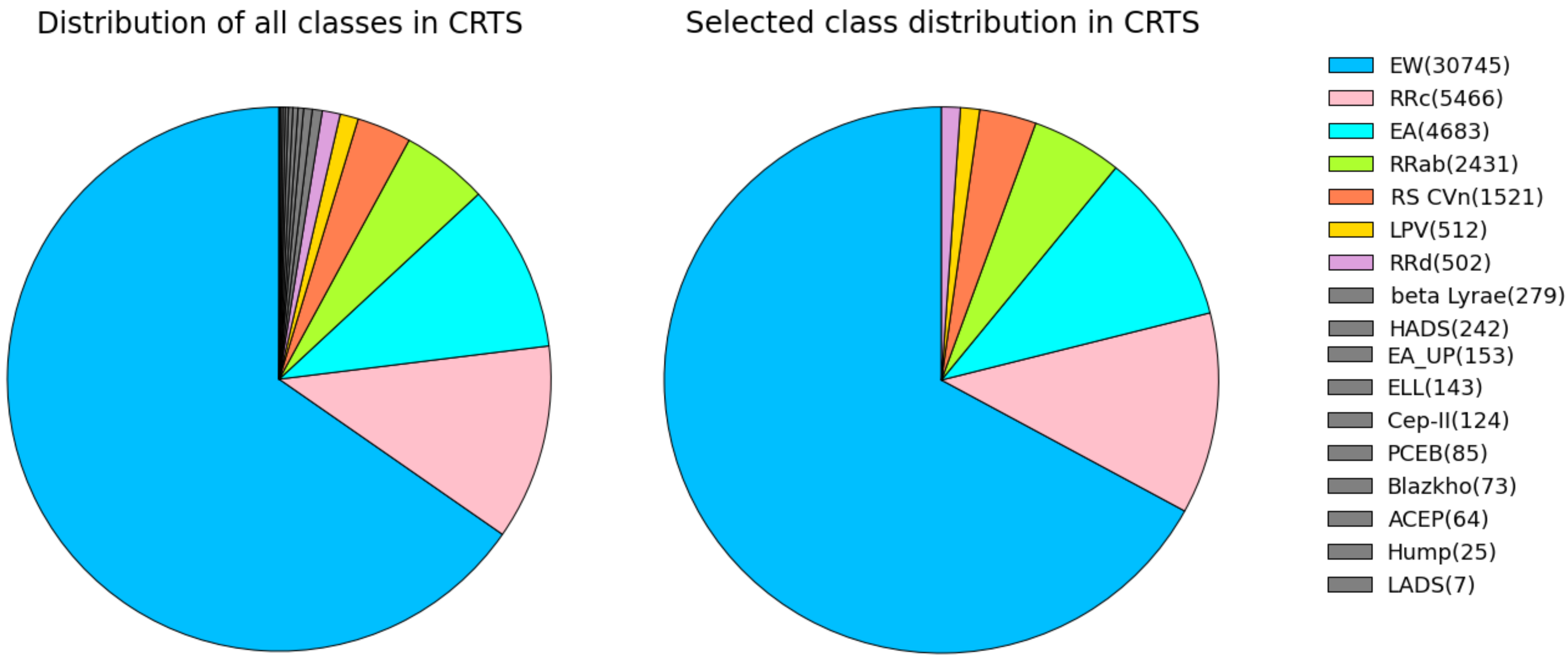


Class / Model 1: basis model with correct period
Class / Model 2: basis model with slightly misspecified period

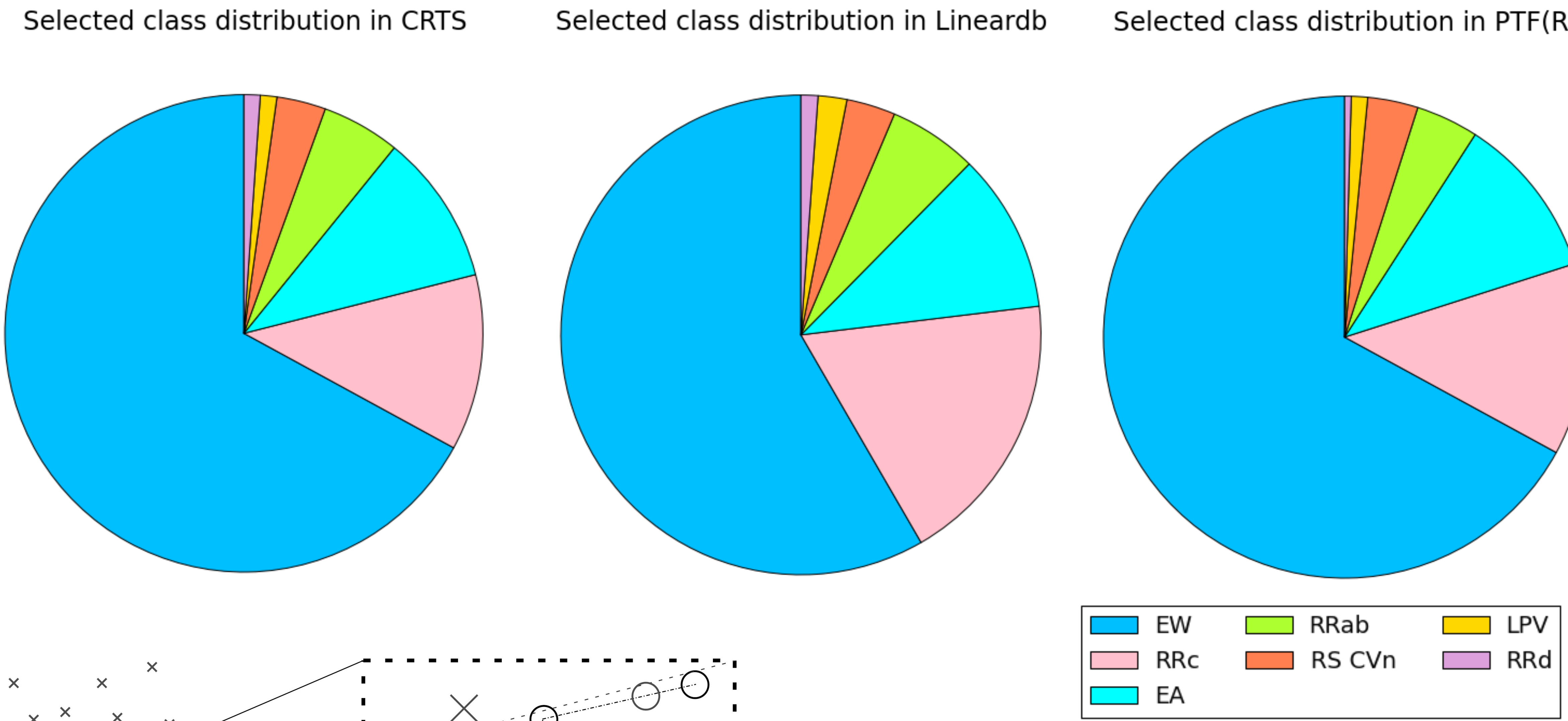
Left: solid green line shows the optimal (posterior mean) time for a new observation in a one day interval indicated by vertical dashed lines. Red and blue curves show current posterior mean fits for models 1 and 2.

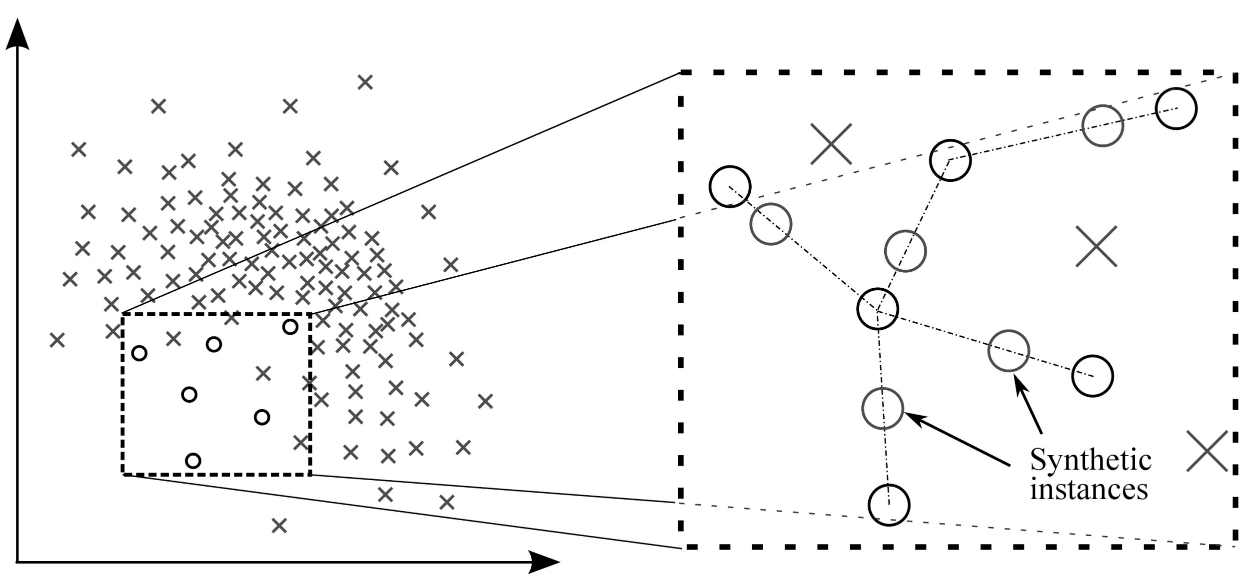
Right: top shows the optimal observation time with the two model means plotted for a single posterior draw of the parameters. Bottom shows the corresponding posterior draw of the separation between the model means

50K Variables from CRTS



Drake et al. 2014





SMOTE and
Sampling with replacement
used to take care of unbalancedness

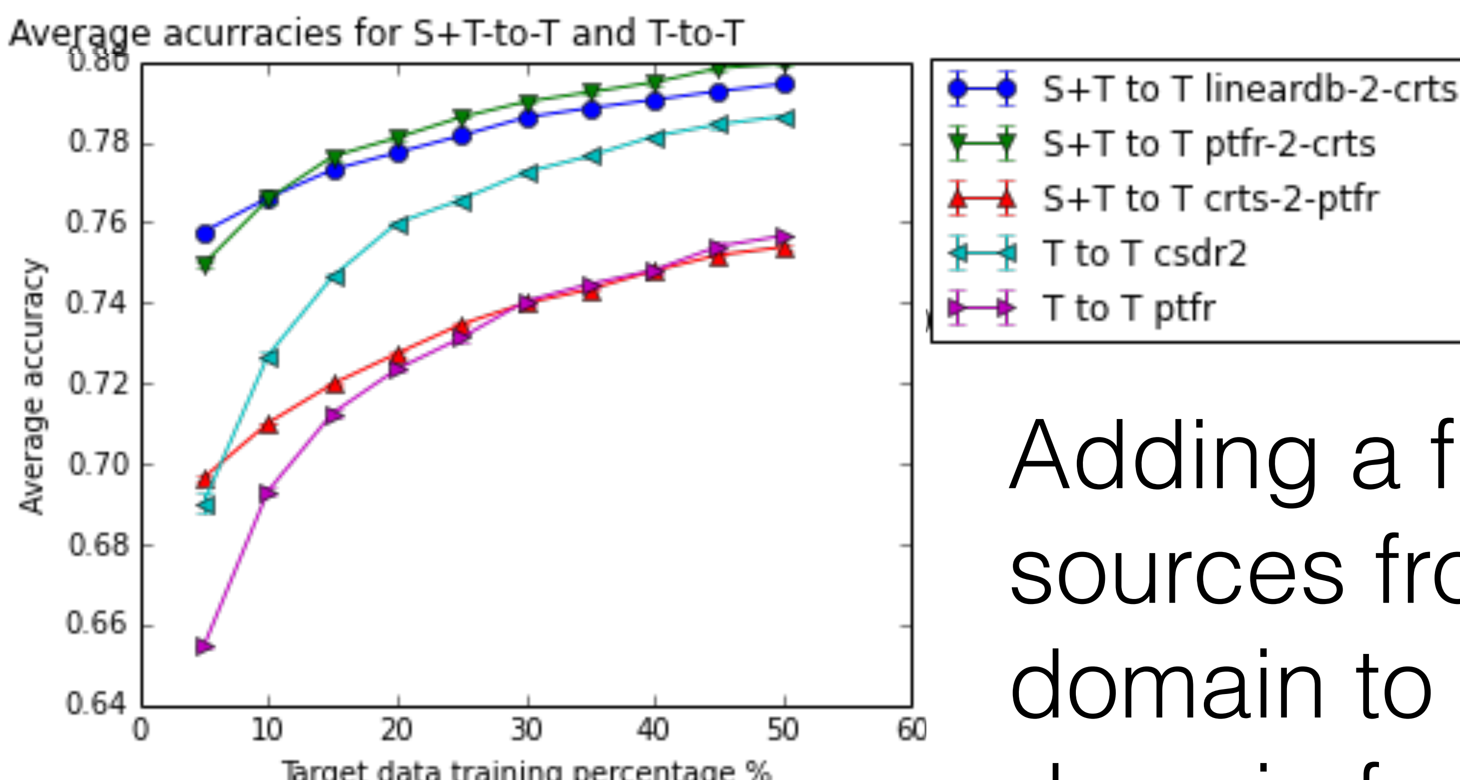
Co-Domain Adaptation

- 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)

$L = D_S \cup D_T^l$

$U = D_T^u$

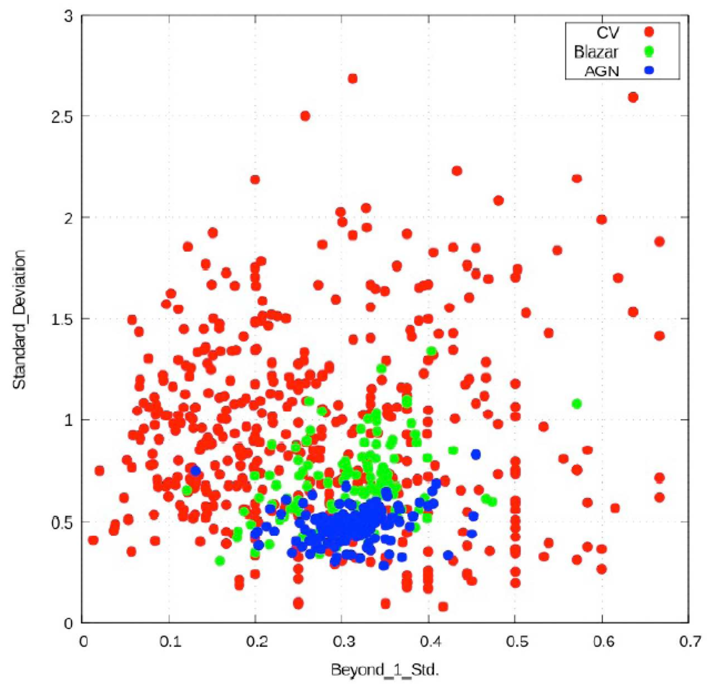
Average Accuracy

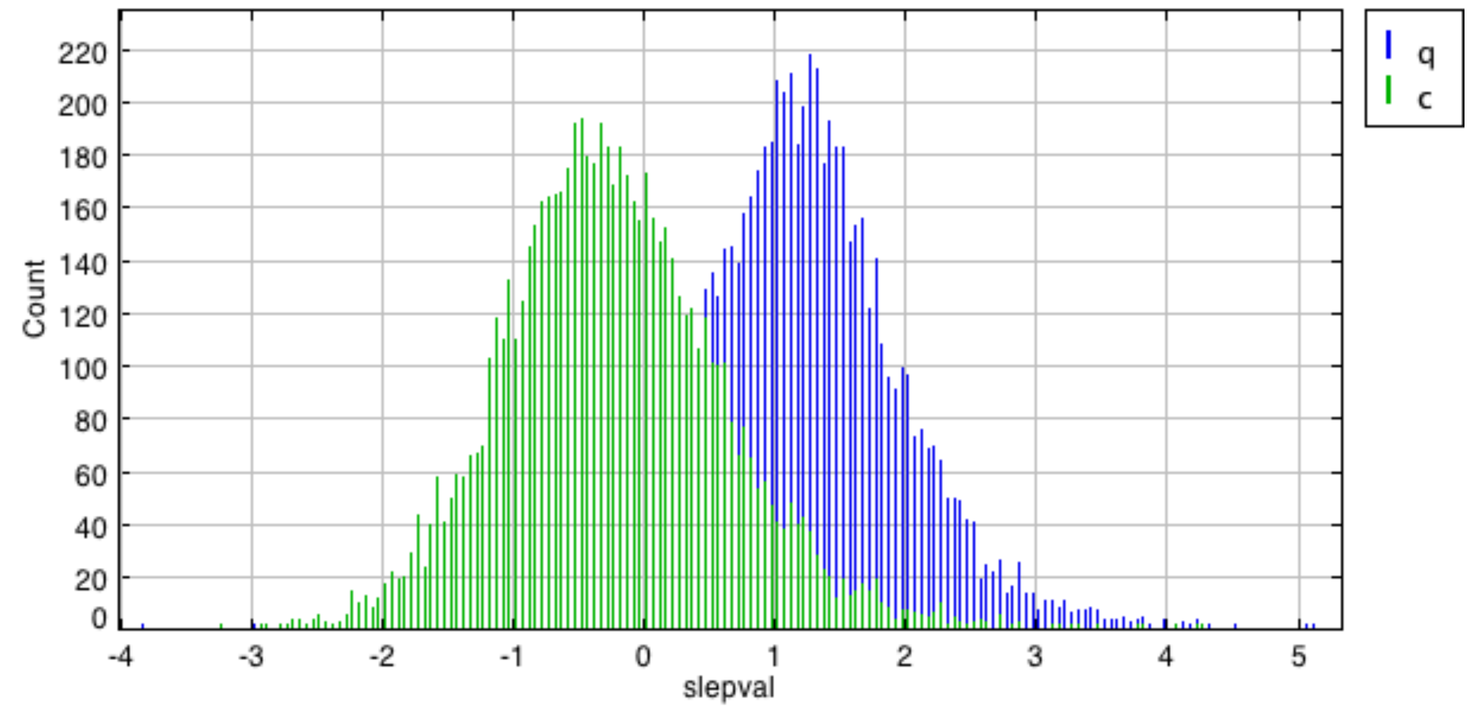


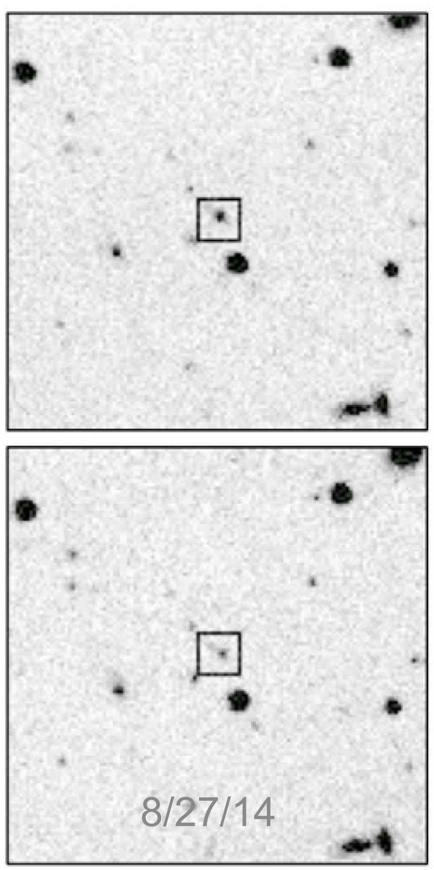
Target Data Training %

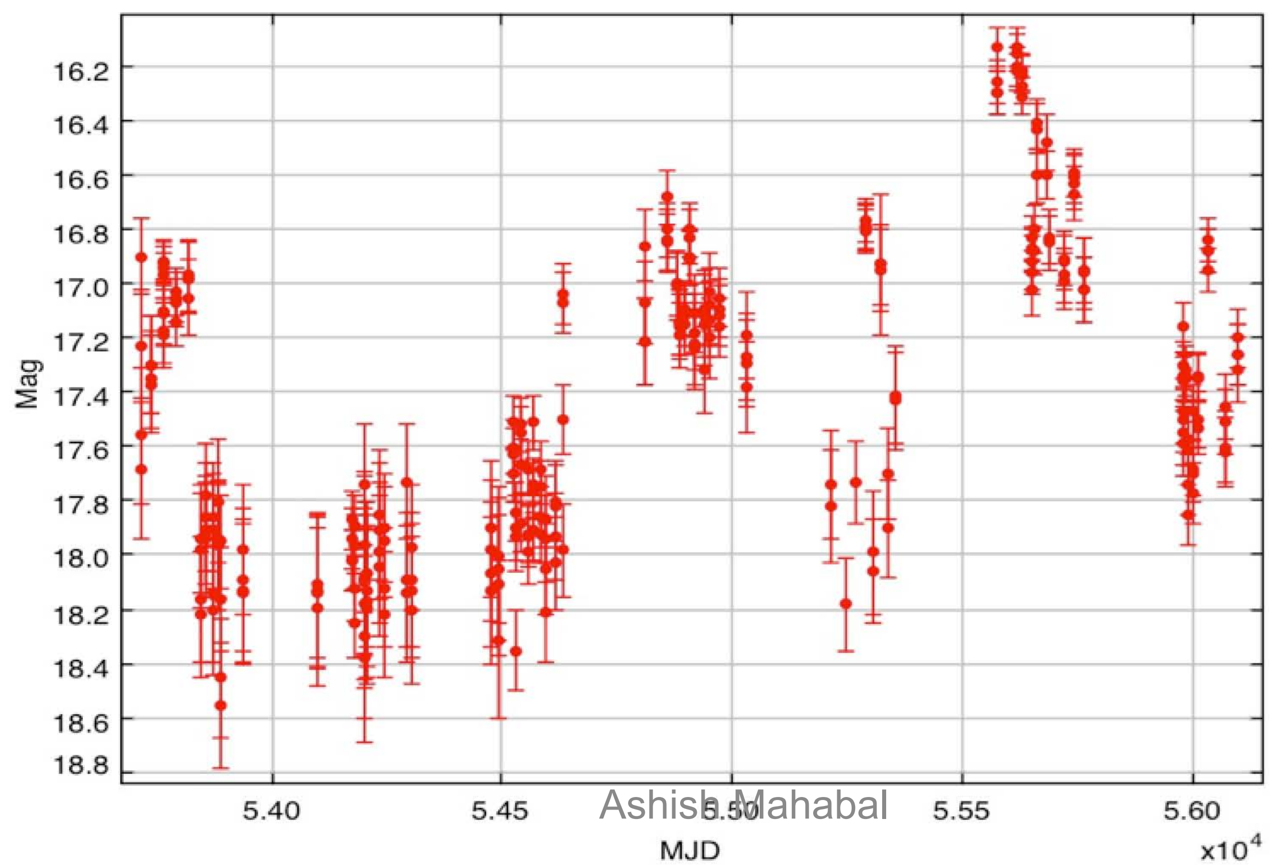
Adding a fraction of sources from the target domain to the source domain for training improves performance

Variability-Based Selection of Quasars





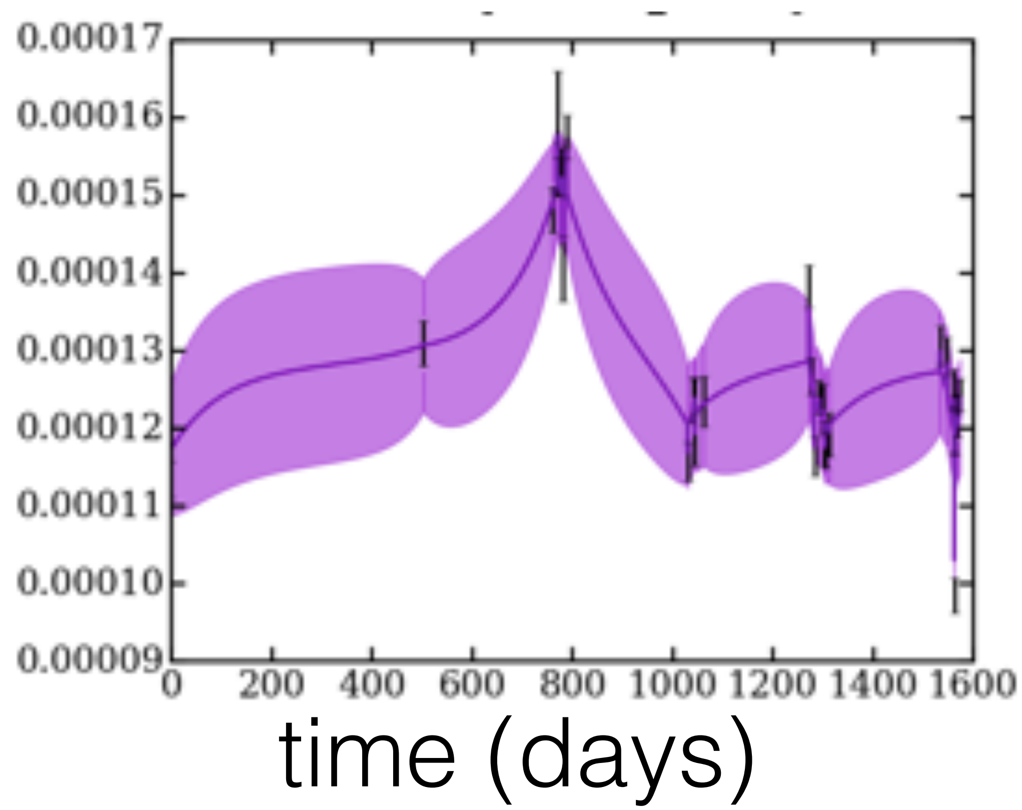




Data mining the light curve variability parameter space to select quasar candidates

(Lead: M. Graham)

Lightcurve decomposition



quasar

CARMA (auto-correlated behavior at various timescales + random disturbances)

CARIMA (non-stationary process)

CARFIMA (long memory process)

Continuous time models are necessary for irregularly sampled data

Summary of challenges

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

Dynamic Time Warping and clustering light-curves

dtwclust (R)

Static data: many clustering methods:

- partitioning (or partitional)
- hierarchical
- density-based
- model-based methods

Time-series are dynamic and pose interesting challenges due to their dimensionality (length, or multivariate i.e. several values changing together) - Aghabozorgi et al. 2015.

dtwclust methods may not work for sparse series, but no one has tried it for astronomy.

.

Possible exercise

Reproduce and extend Faraway et al. 2014

http://people.bath.ac.uk/jjf23/modlc/

Exploratory analysis
Gaussian Process regression
Derived Measures
Variable importance

Jupyter notebook prepared by Melissa Hayes-Gehrke for GRWOTH

https://nbviewer.jupyter.org/url/growth.caltech.edu/quick-view/light_curve_solutions.ipynb

PLAsTiCC data challenge

https://www.kaggle.com/c/PLAsTiCC-2018