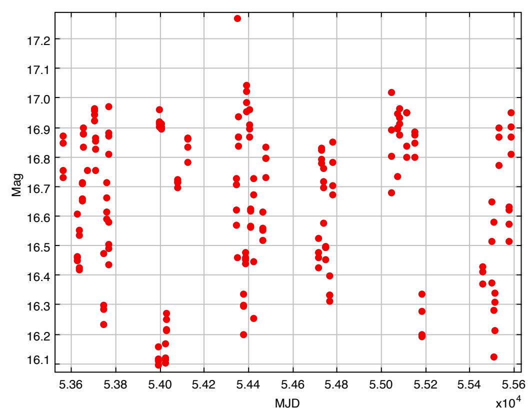
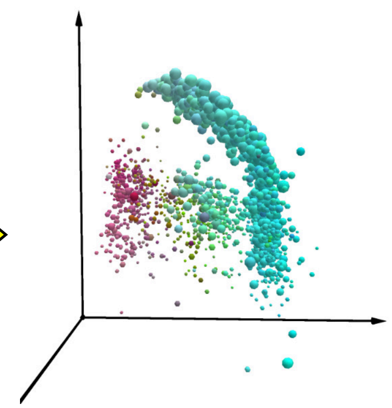
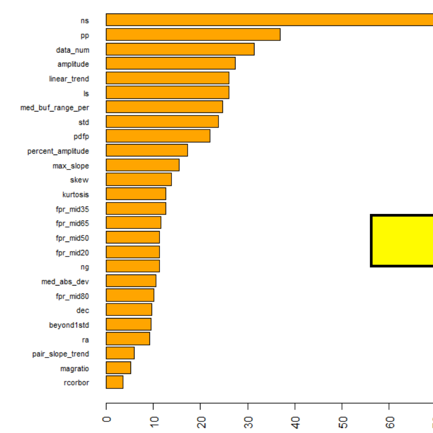
Feature based Light Curve Classification



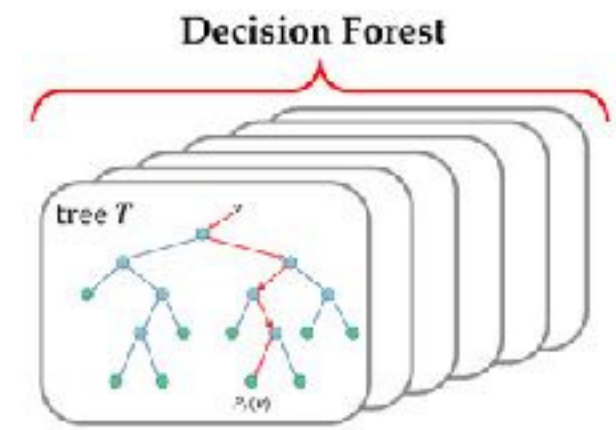
Light curves



Feature vectors



Dimensionality Reduction



Classification

Ashish Mahabal
Center for Data Driven Discovery, Caltech

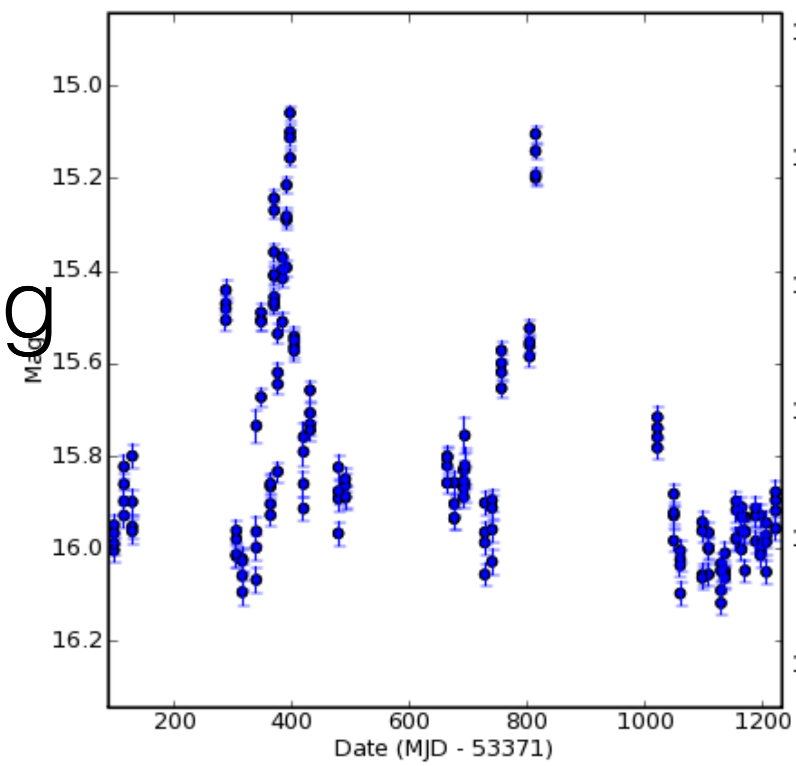
La Serena School for Data Science
La Serena, 24 Aug 2018

Outline

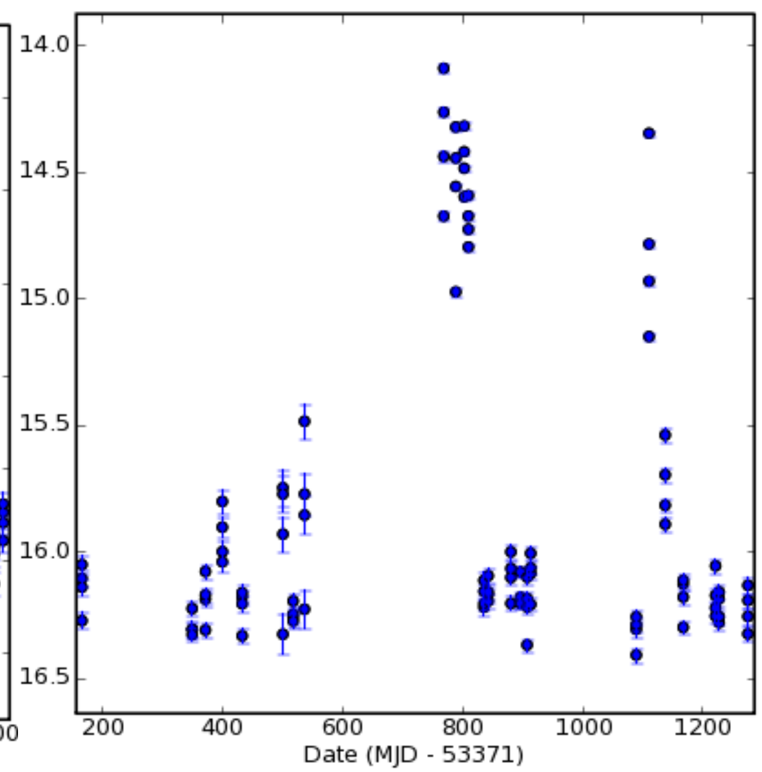
- Surveys and light curves
- Need for classification
- Statistical features
- Classification
- [Examples/Exercises]

Time Series aka light-curves we will encounter

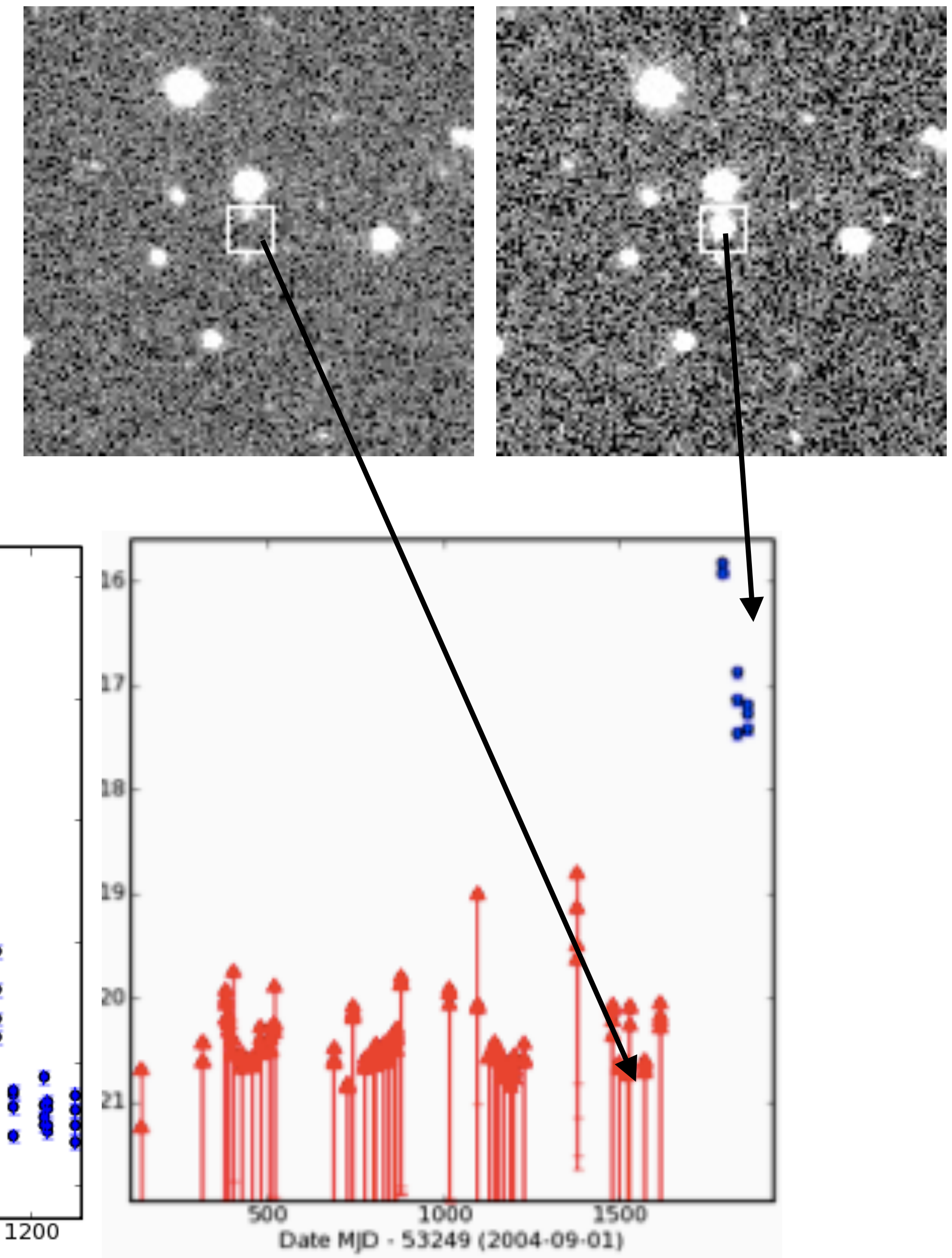
Blazar PKS0823+033



CV 111545+425822



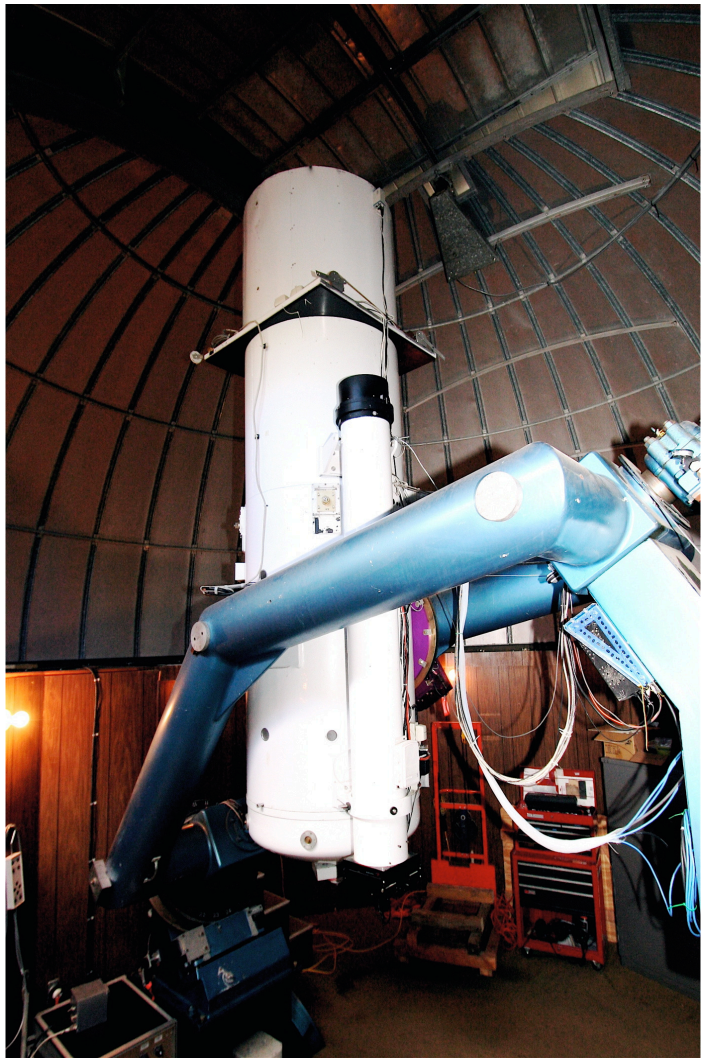
mag



Supernova

Time (1000+ days)

magnitude is logarithmic, inversely scaled (flux)



1m class telescopes
~20 mag

Open filter
~14 years
500M light-curves

23000 sq. deg
(moon ~ 0.25 sq deg

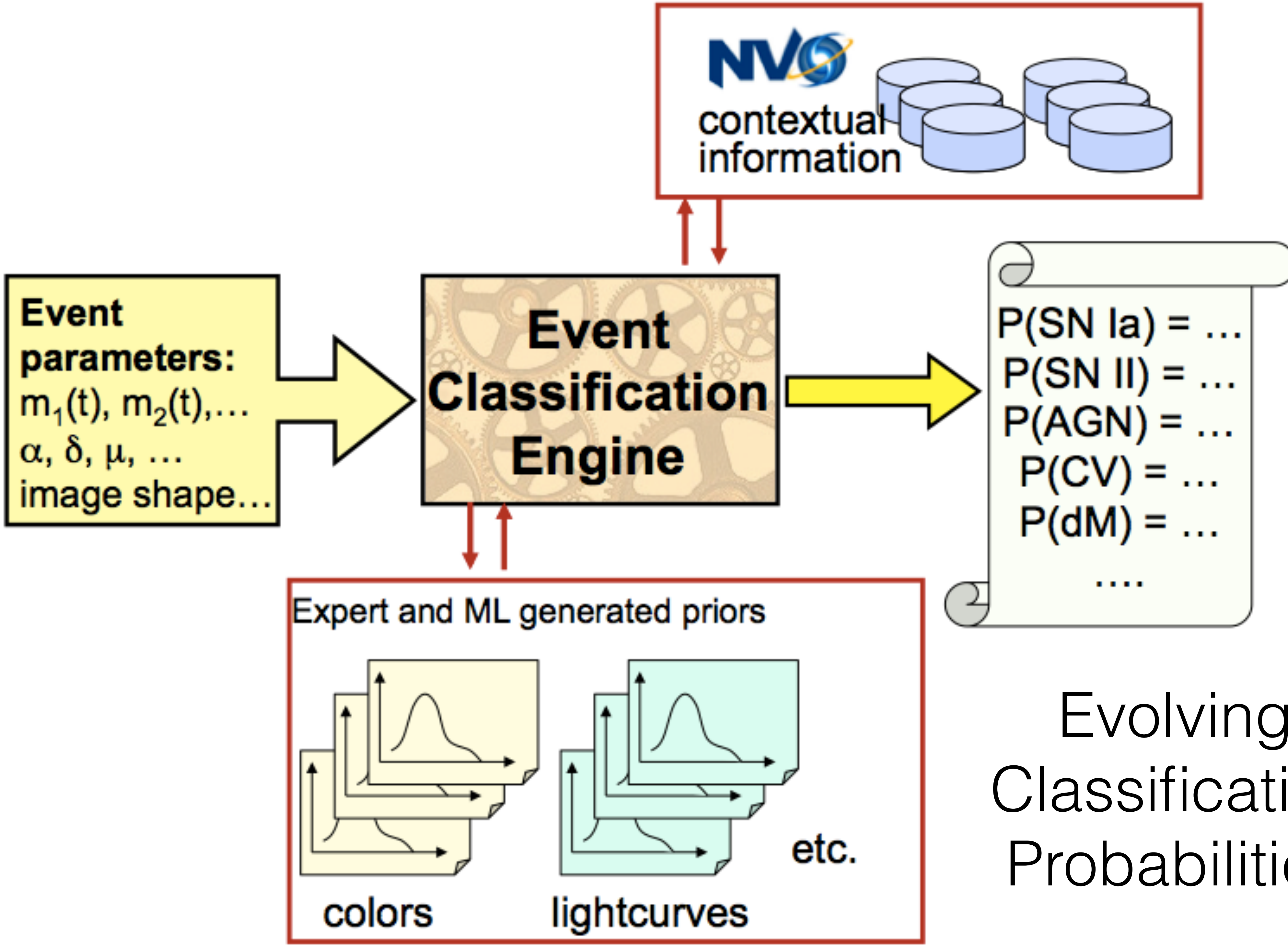
CRTS

Transient
Searches

~200 pointings
30 seconds each

CSS PI: Eric Christensen
CRTS PIs: George Djorgovski
and Andrew Drake

A few years ago …



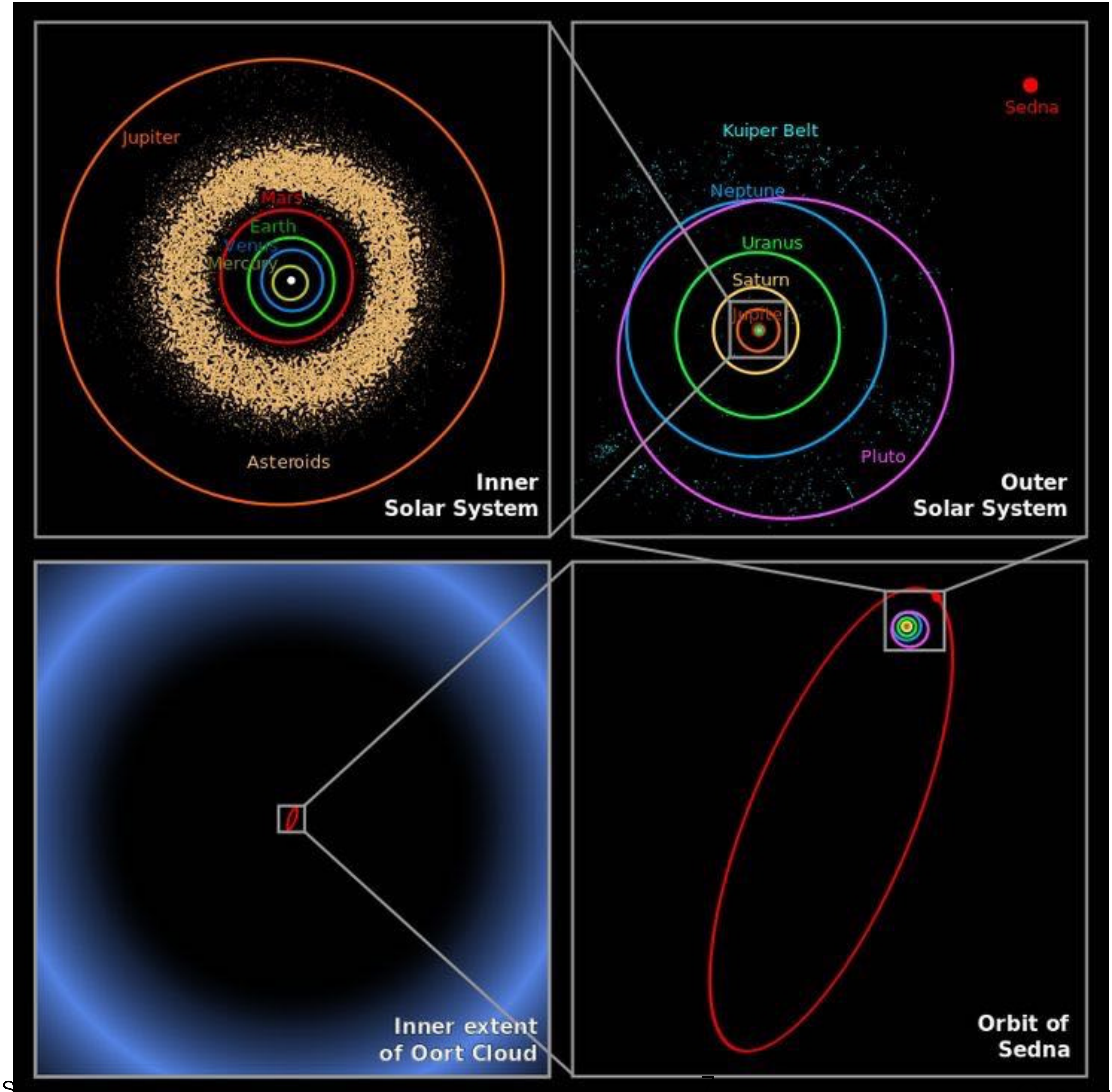
Evolving Classification Probabilities

Broad classes in astronomy

Aim:

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

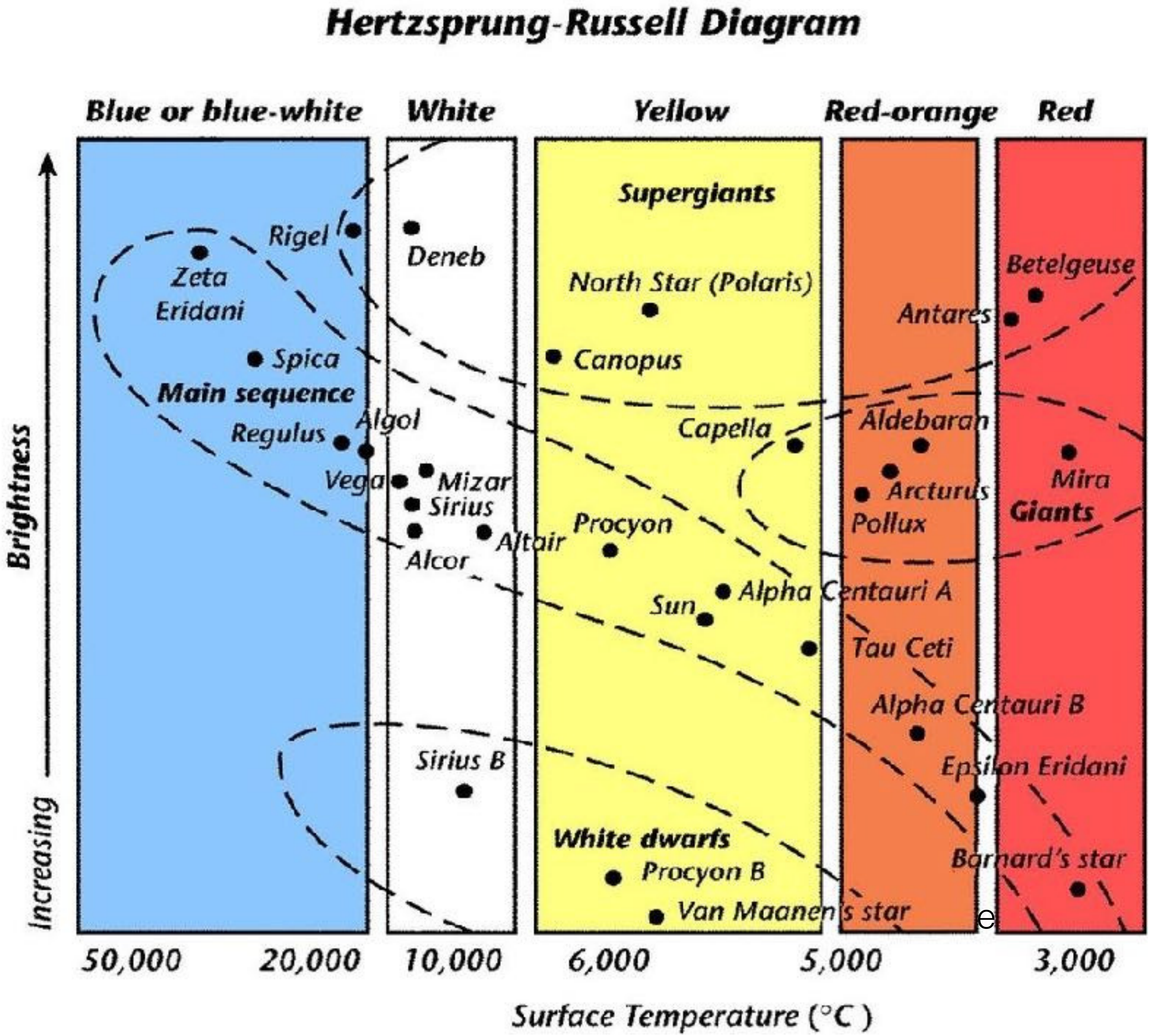
Solar System



Moving objects

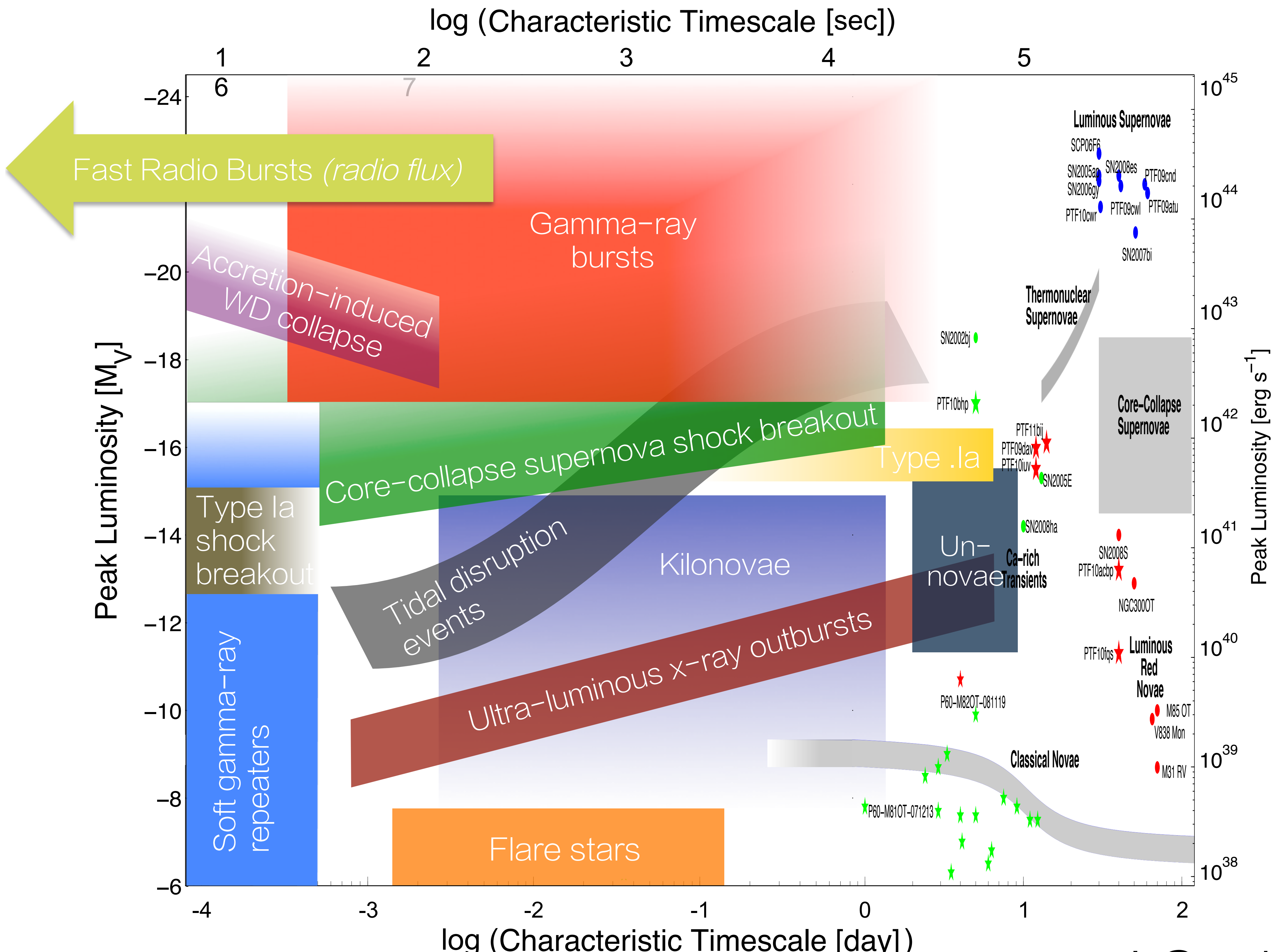
As

Credit: web.gps.caltech.edu



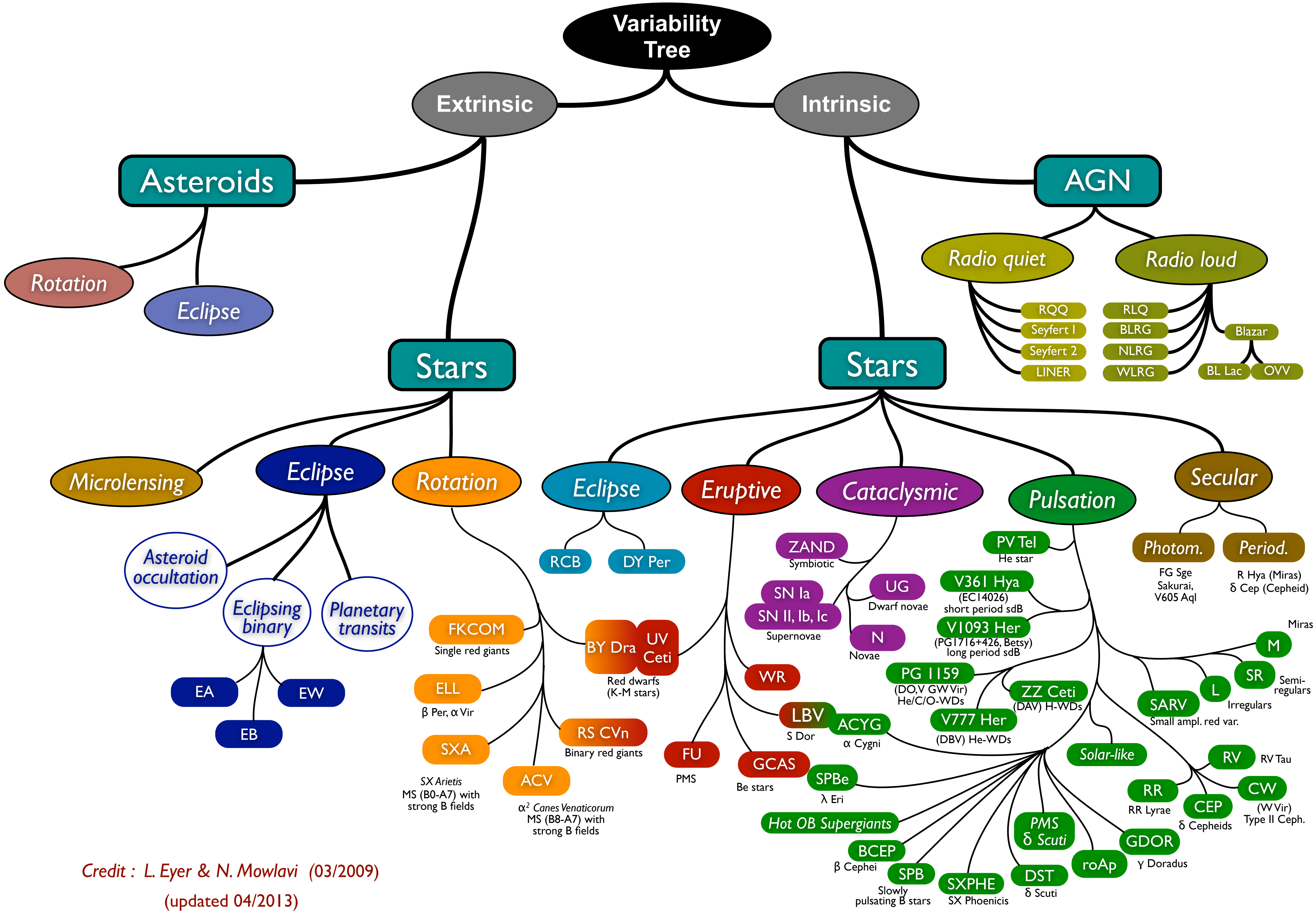
http://www.huntsville-isd.org/

Transients (mostly extra-galactic)



J Cooke

Variability tree: Many nodes have further subdivisions

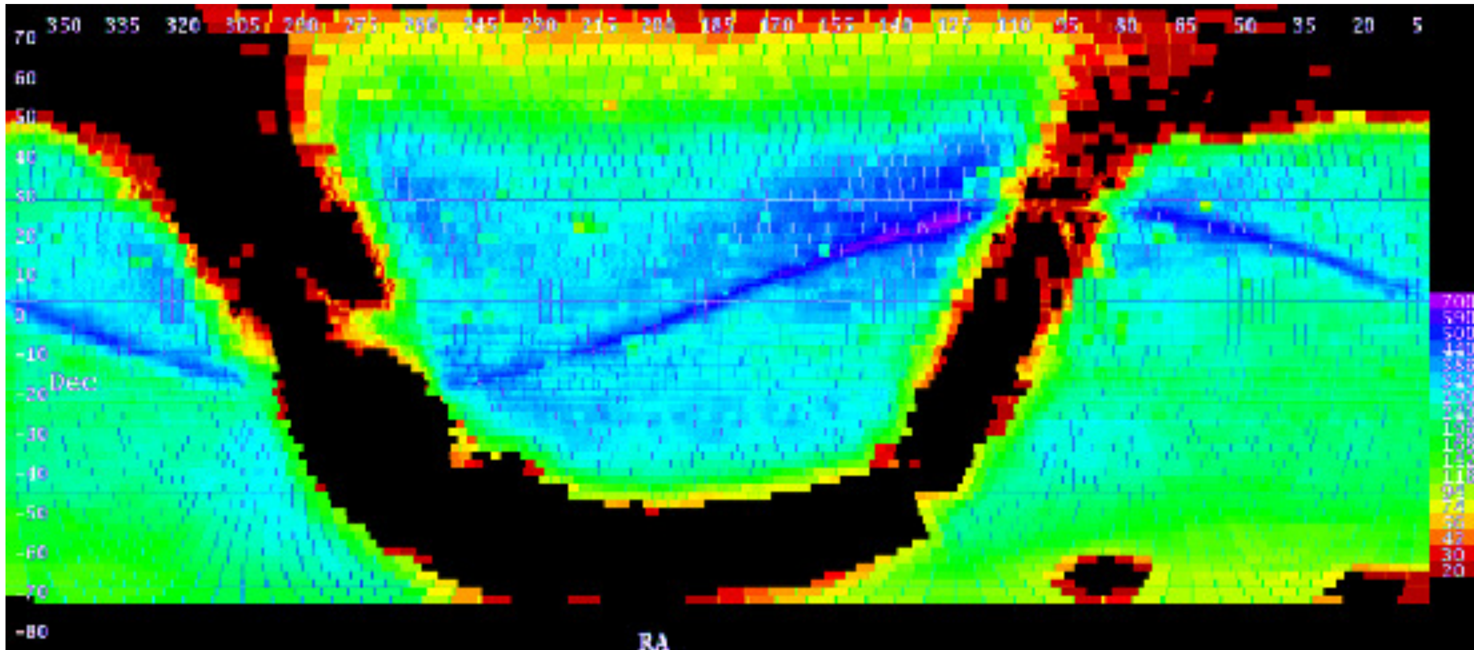


Credit : L. Eyer & N. Mowlavi (03/2009)

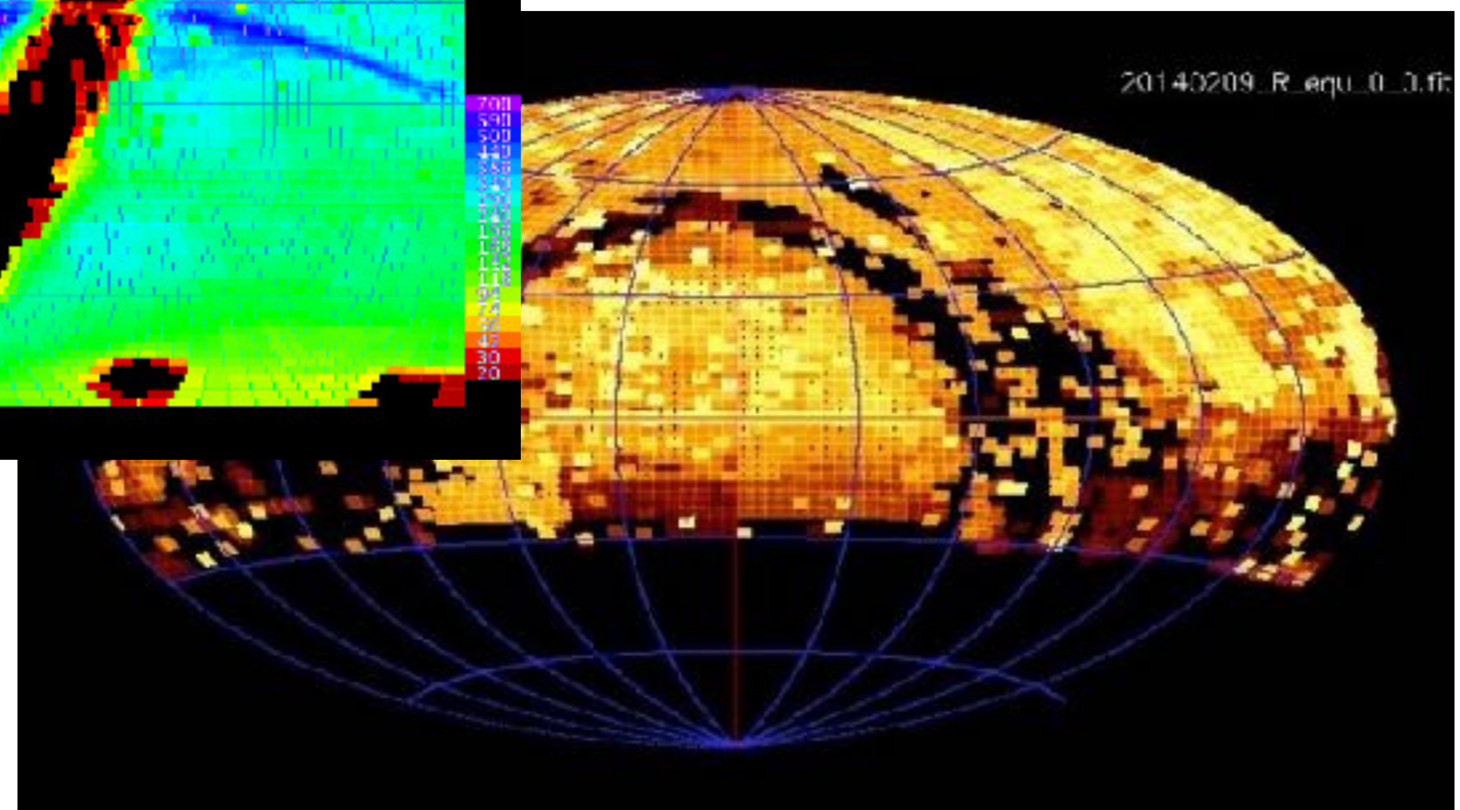
(updated 04/2013)

From snapshots to (slow) movies of the sky

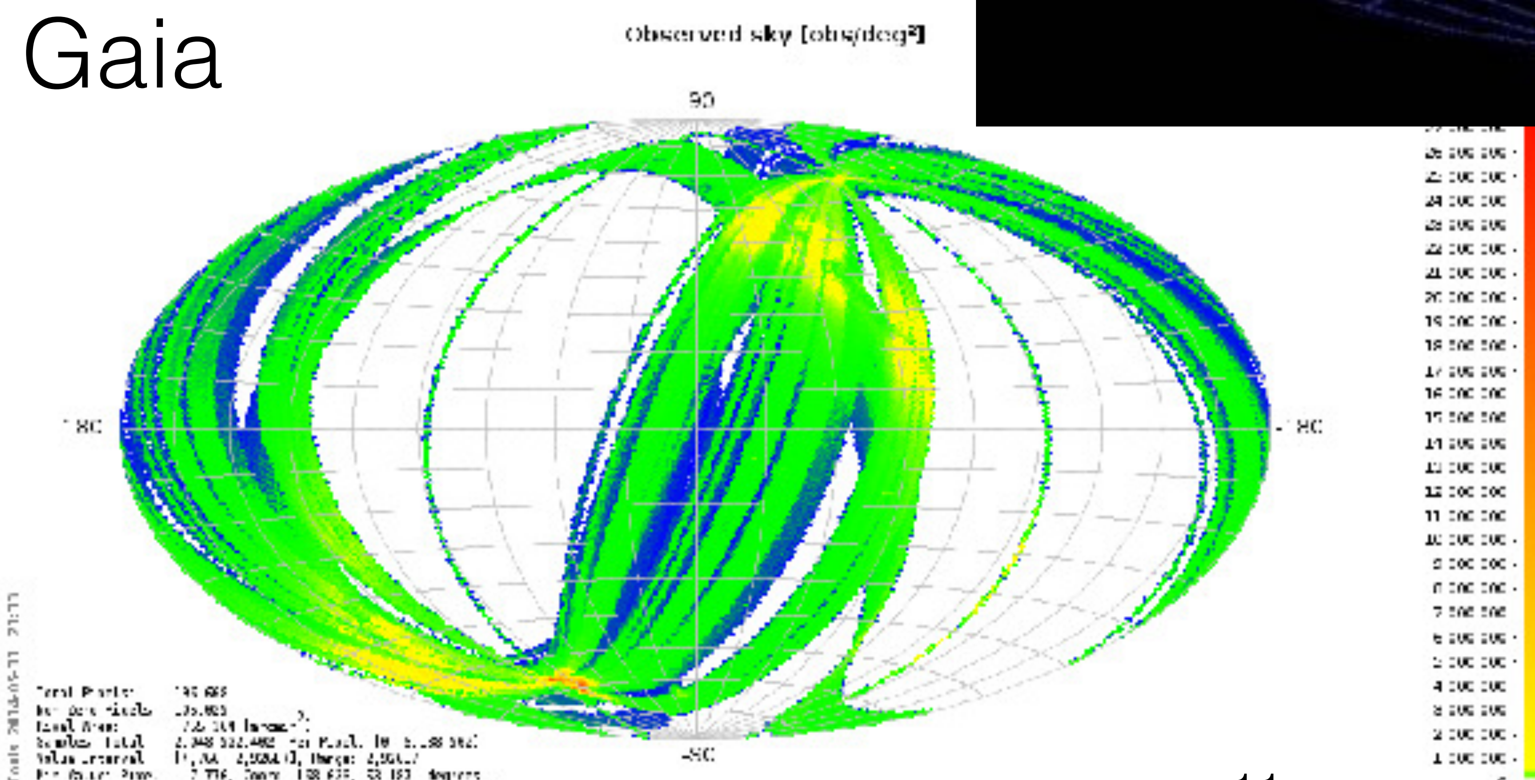
CRTS

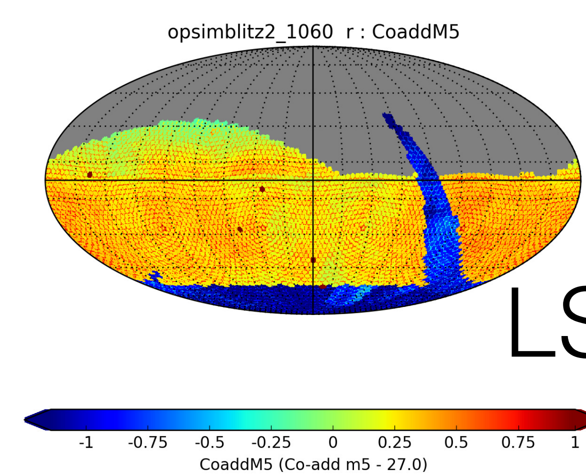


PTF



Gaia





LSST

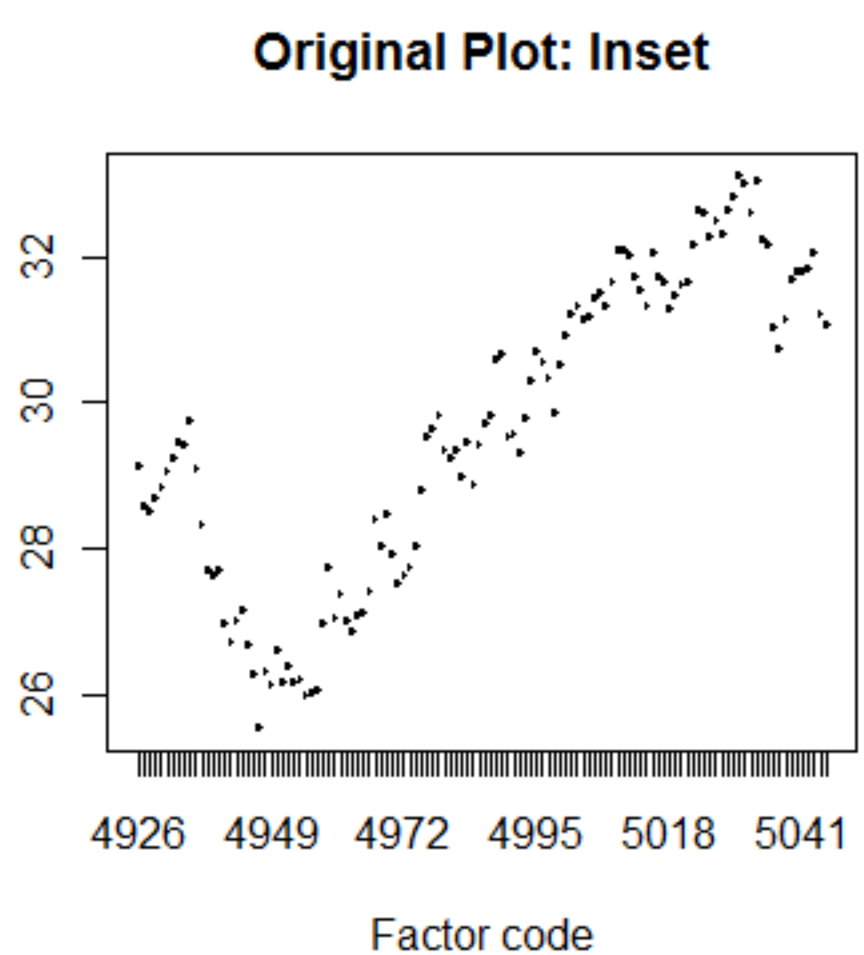
What do survey's do?

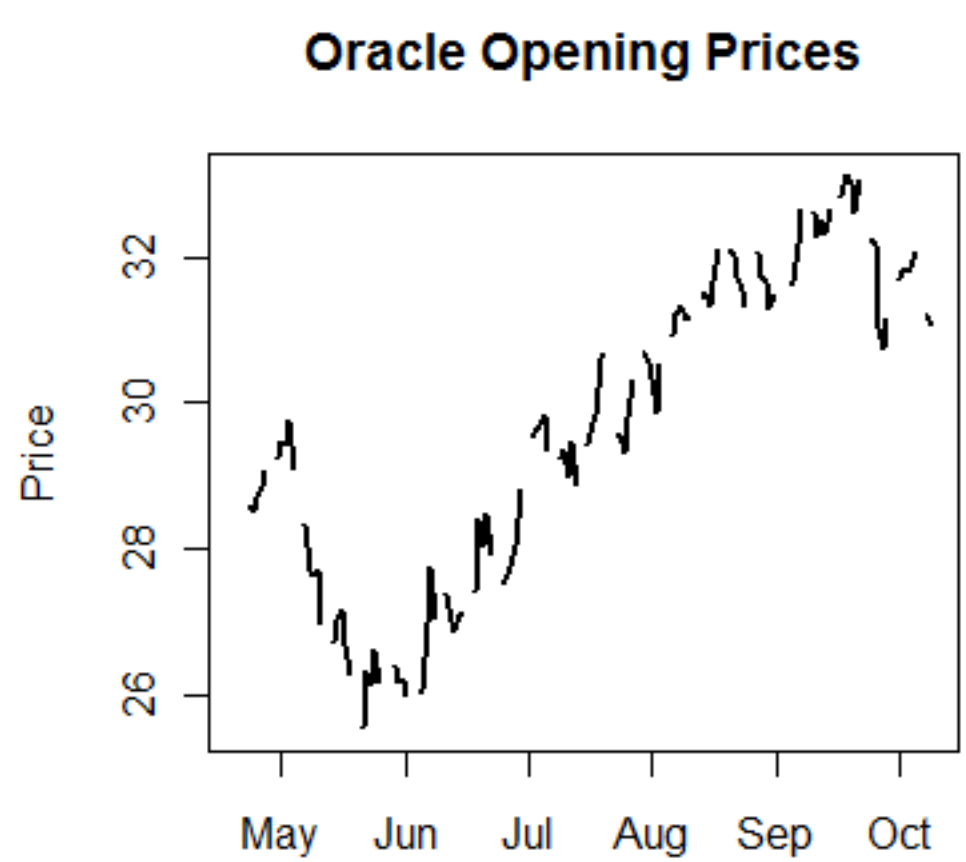

- Pick low-hanging fruit
  - select best objects, easy science
  - get spectroscopy
- That does push the envelope
  - but also leaves gaps

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

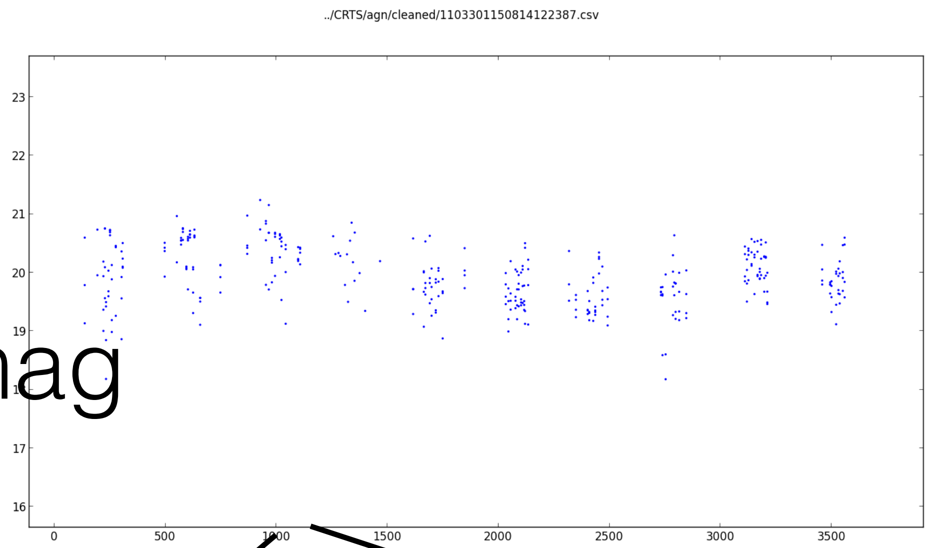




stackexchange

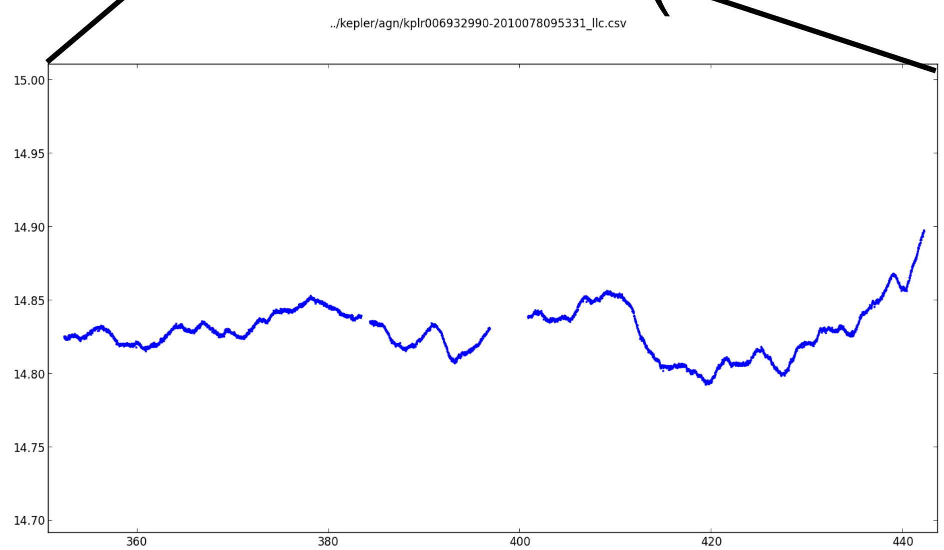
CRTS

Kepler - small area non-sparse

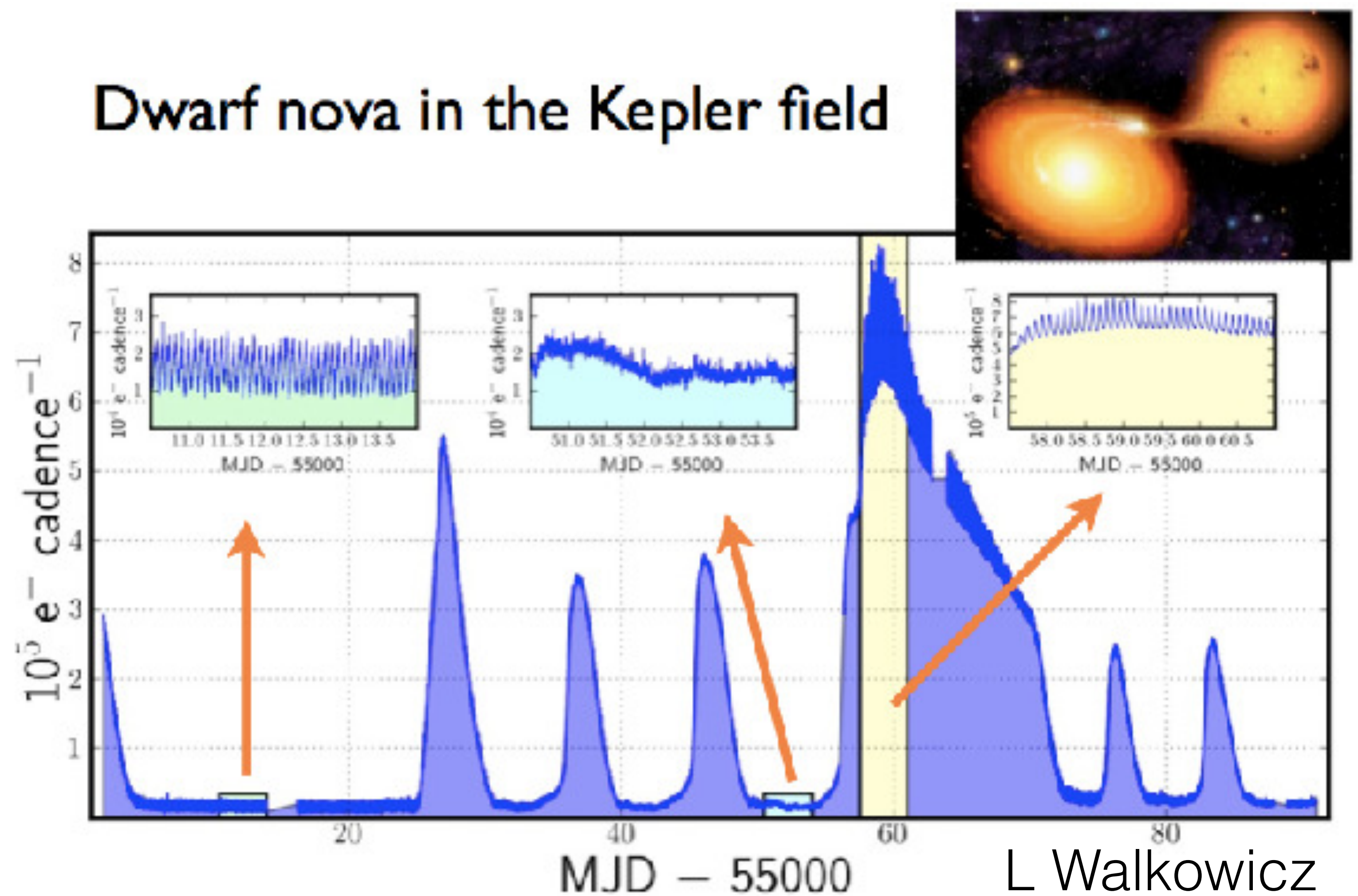


mag

Time (3000+ days)



~100 days



L Walkowicz

Time | Variable | Error

mjd | mag | magerr

Kepler - small area non-sparse

modified JD

JD = days since

12 noon 1 Jan -4712

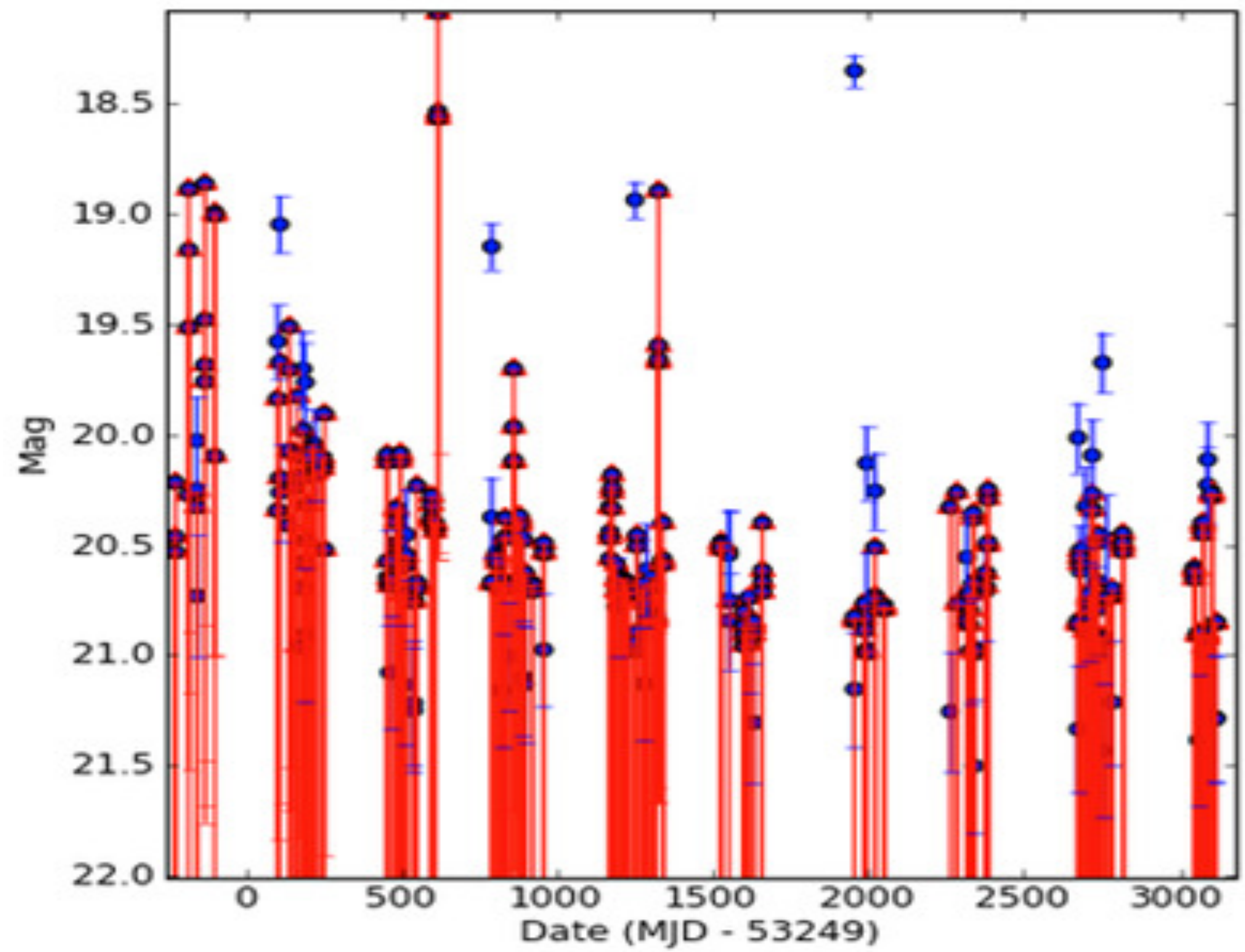
MJD = JD - 2400000.5

Typical time-series in astronomy

- DPOSS - large area, serendipitous overlap
- Kepler - small area non-sparse
- **CRTS - open filter, lumpy cadence for asteroids**
- **PTF/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

CRTS variables

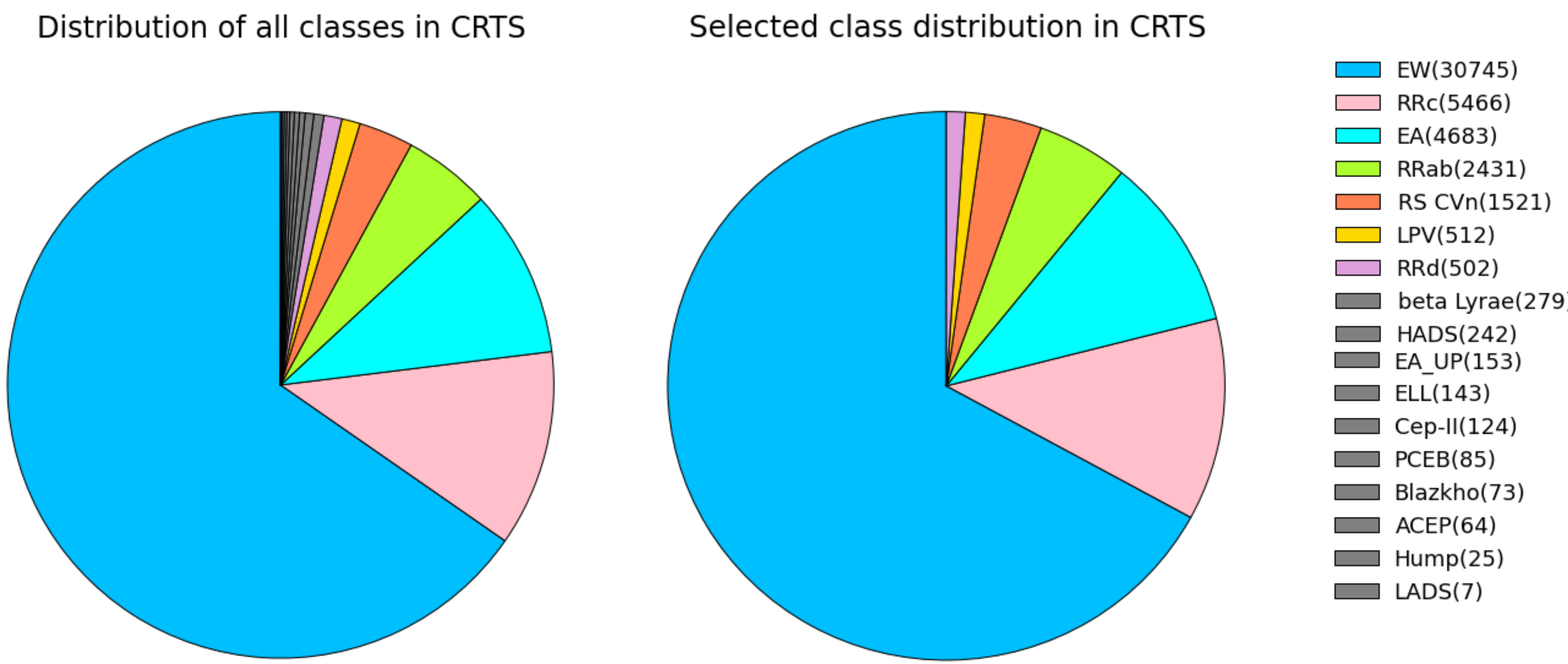
- 150M sources from a few thousand “fields”
- ~5.5M variables after filtering using per field J
- ~50K periodic (LS False Alarm Probability < 10^-5; M_t thresholds)

Drake et al. 2014

  - 15 classes

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

50K Variables from CRTS



Drake et al. 2014

Over to part 1 of notebook

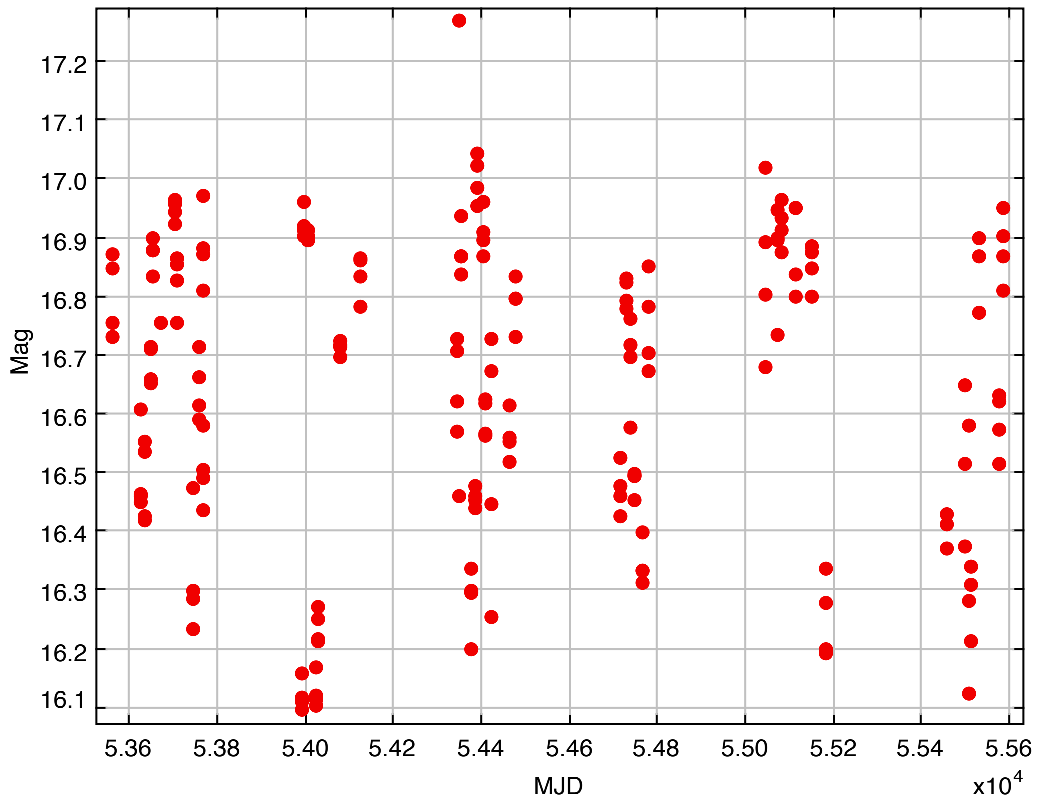
What can we do with light-curves?

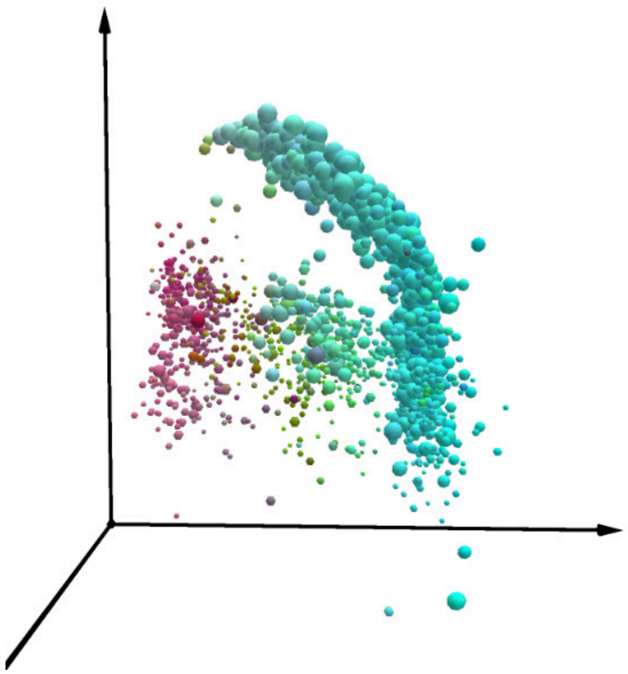
dreamstime.com

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

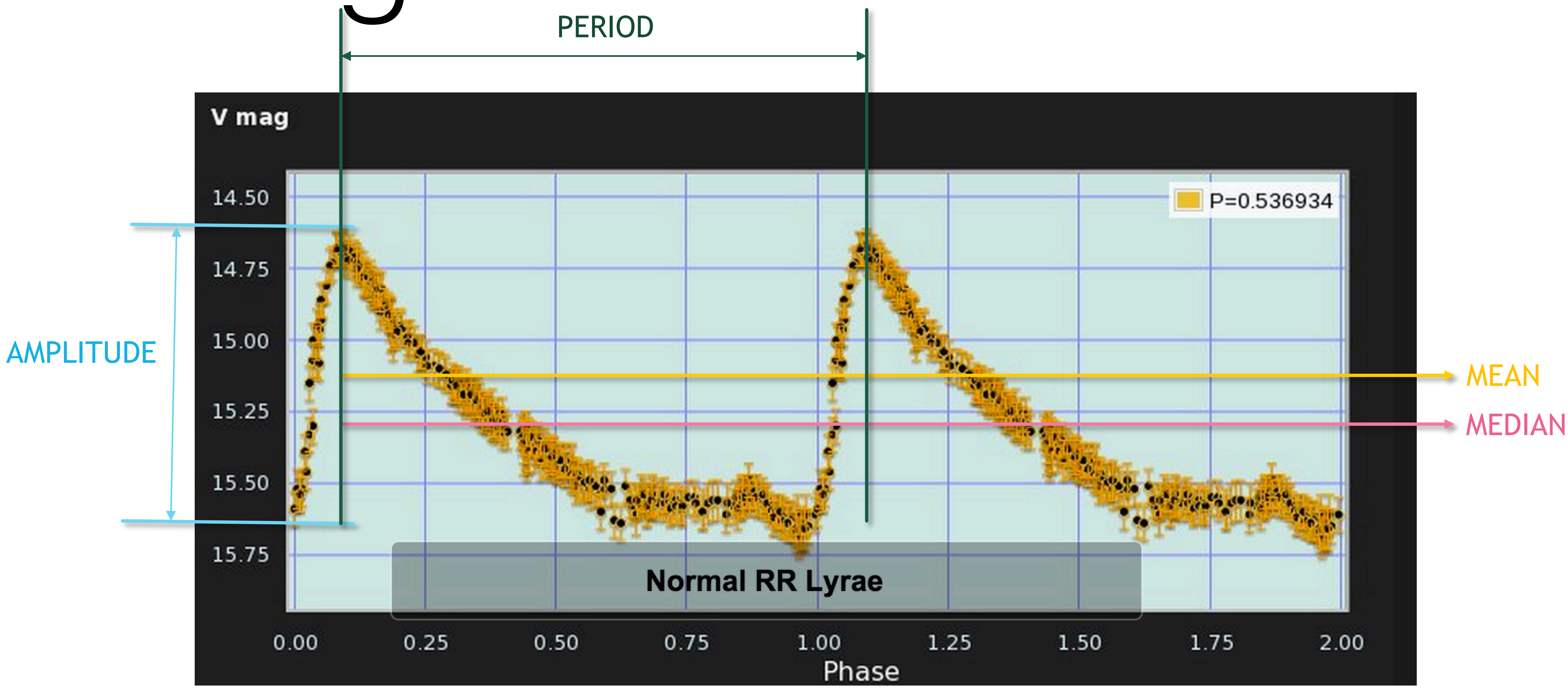
Statistical features

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





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

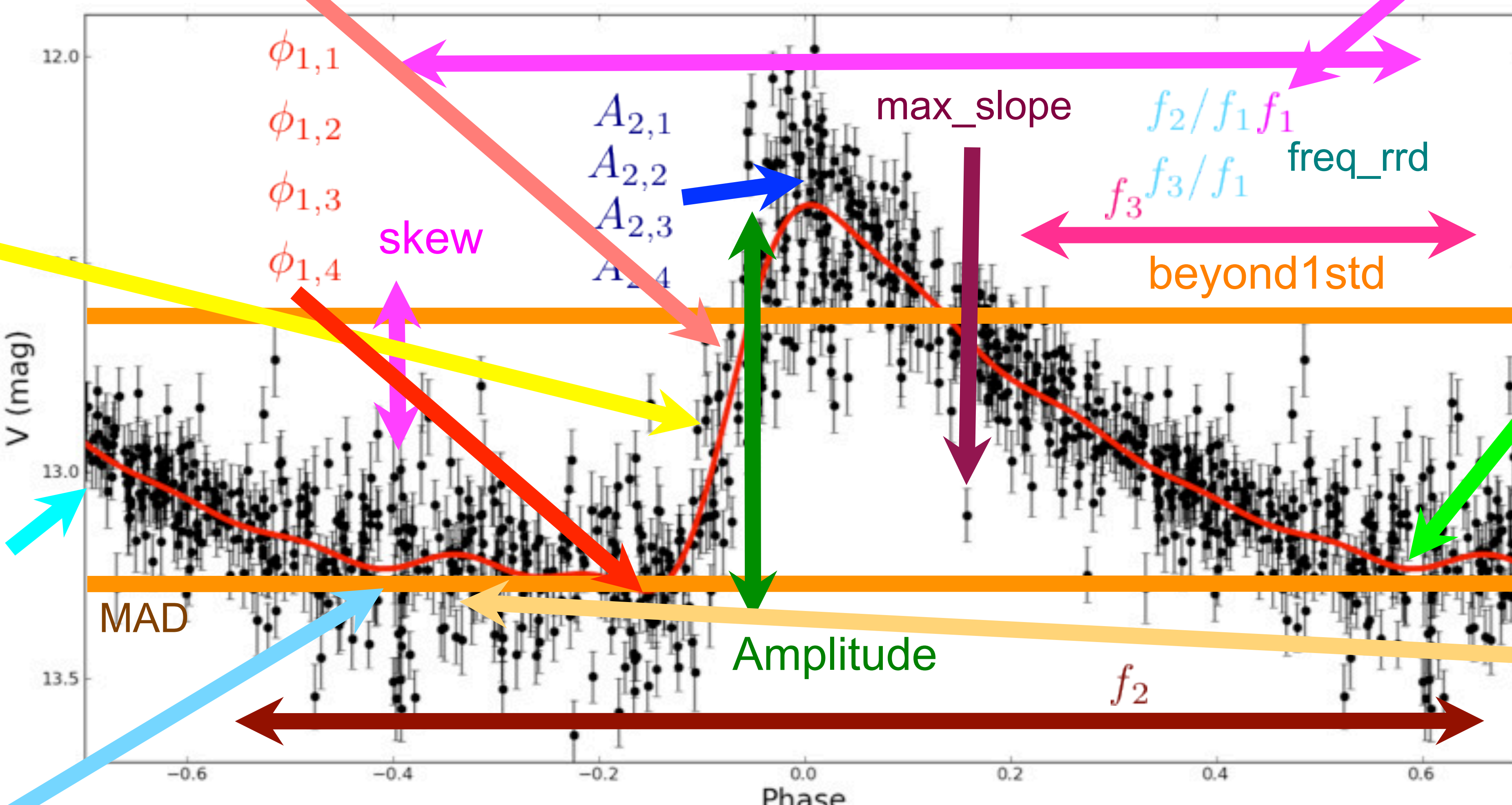
amplitude

Short name	Data type	Summary
amplitude	float	0.5 * (mag_{max} - mag_{min})
beyond1std	float	p(|(mag - <mag>)| > σ)
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

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

Combined for thresholding

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-

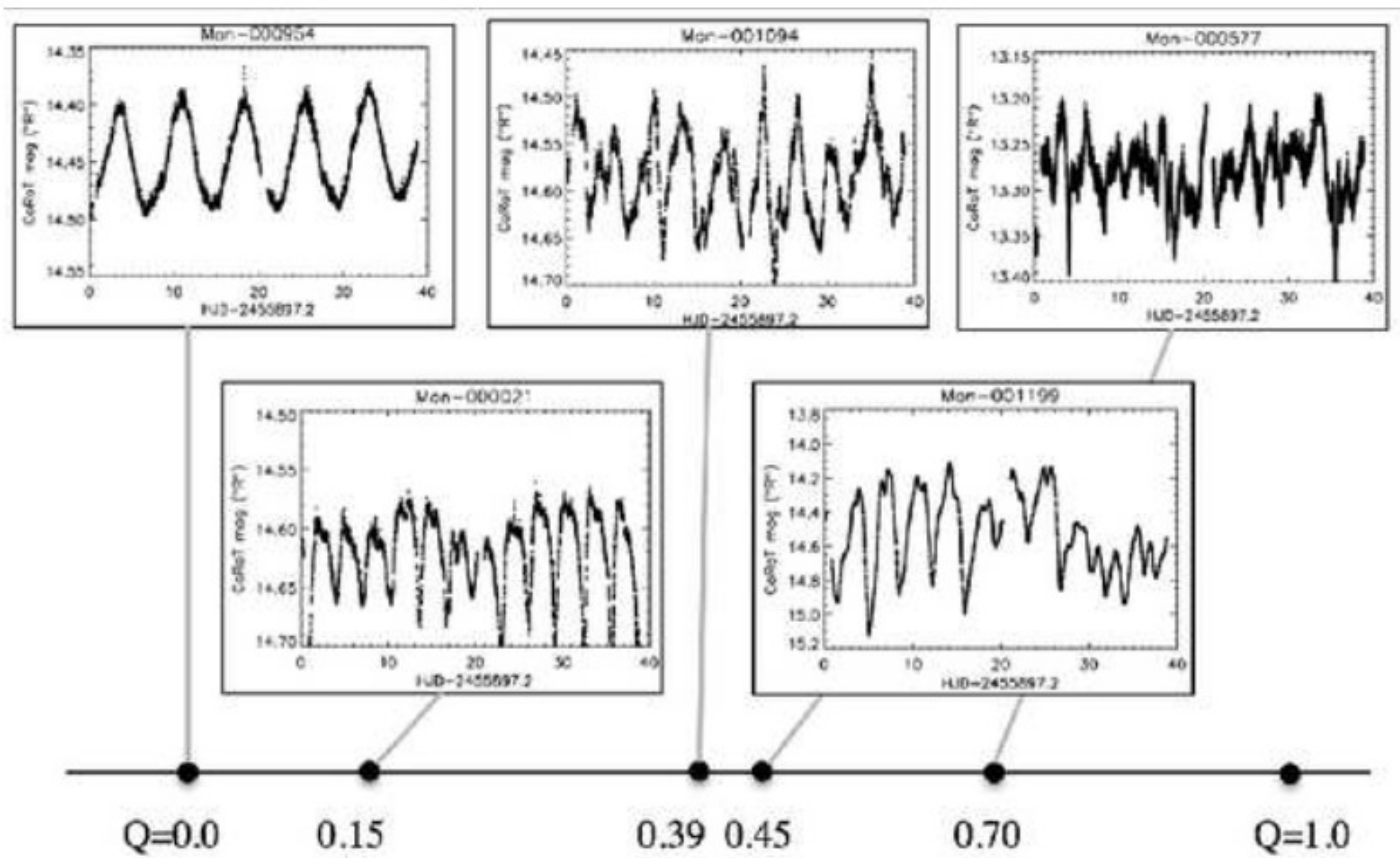


FIG. 29.— *CoRoT* light curves with representative values of the Q parameter, ranging from periodic (Q=0–0.15) to quasi-periodic (Q=0.15–0.5), to aperiodic $Q > 0.5$.

Cody et al. 2014

M: Bursters and dippers

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

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

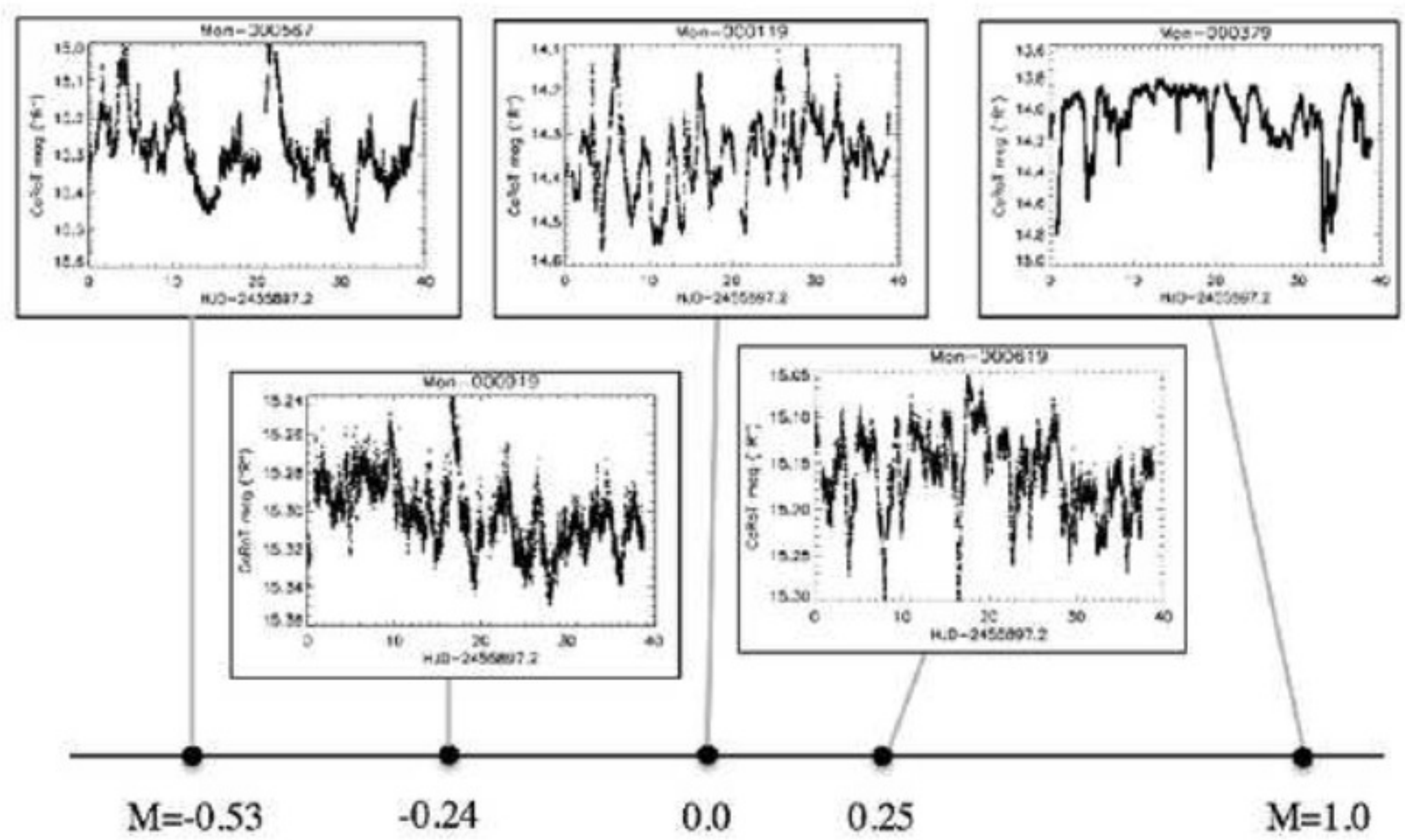


FIG. 30.— *CoRoT* light curves with representative values of the M parameter, ranging from bursting ($M < -0.25$) to symmetric (M=-0.25–0.25), to dipping $M > 0.25$.

Cody et al. 2014

Q-M plane

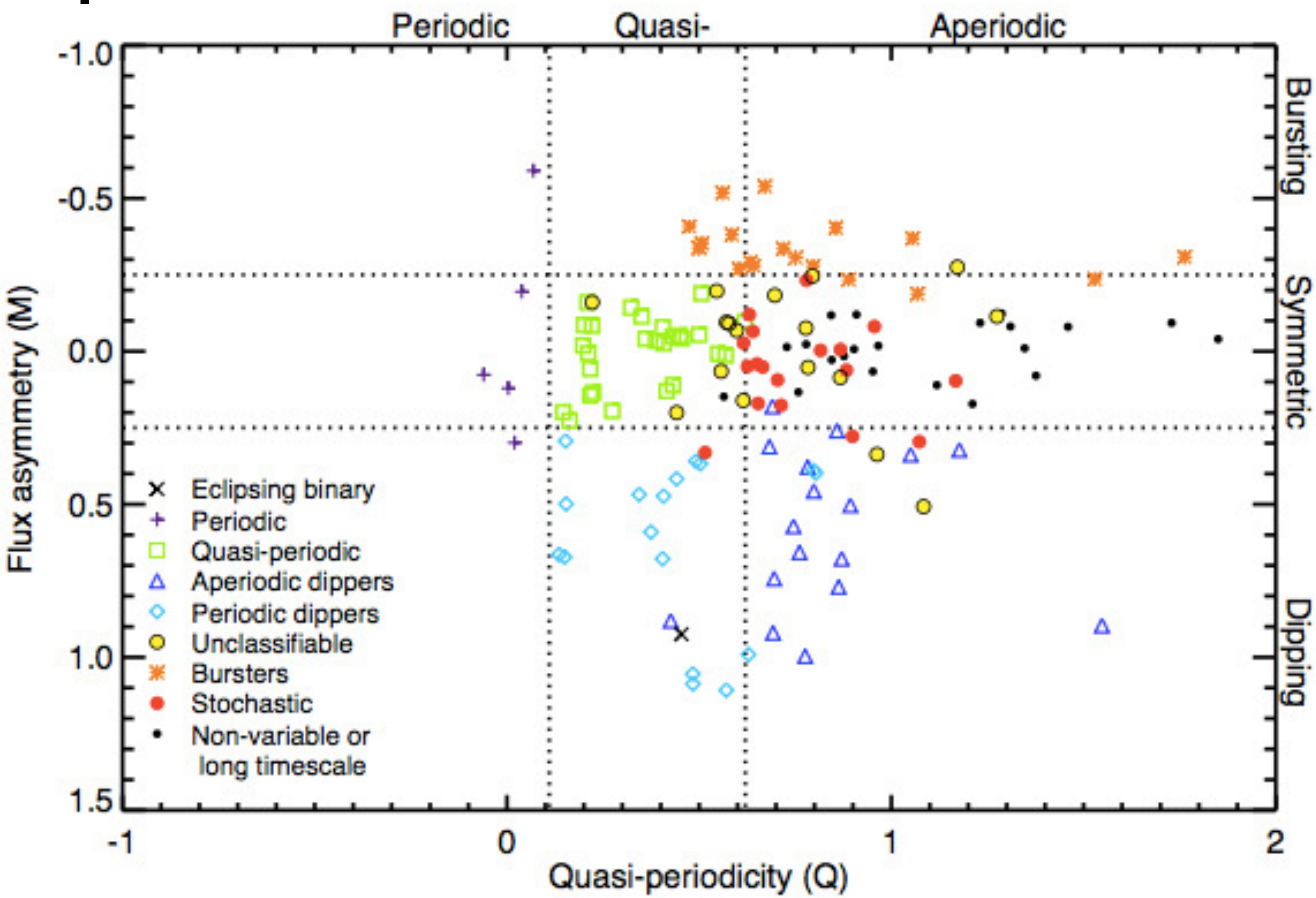


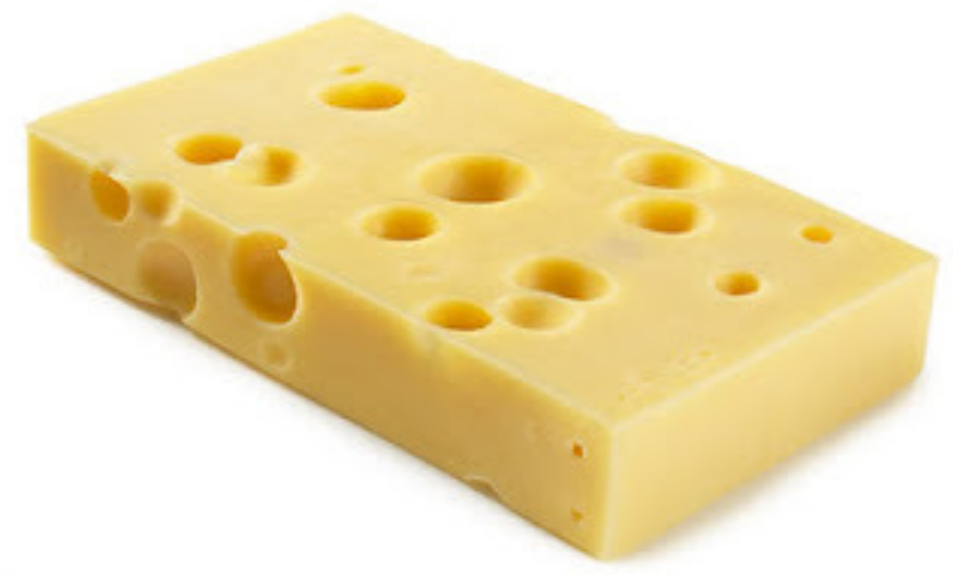
FIG. 31.— *Top*: Light curve morphology classes, as divided by the quasi-periodicity (Q) and flux asymmetry (M) parameters for optical light curves from *CoRoT* in our disk-bearing sample. Color coding indicates the variability classification chosen by eye, before statistical assessment. The eclipsing 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

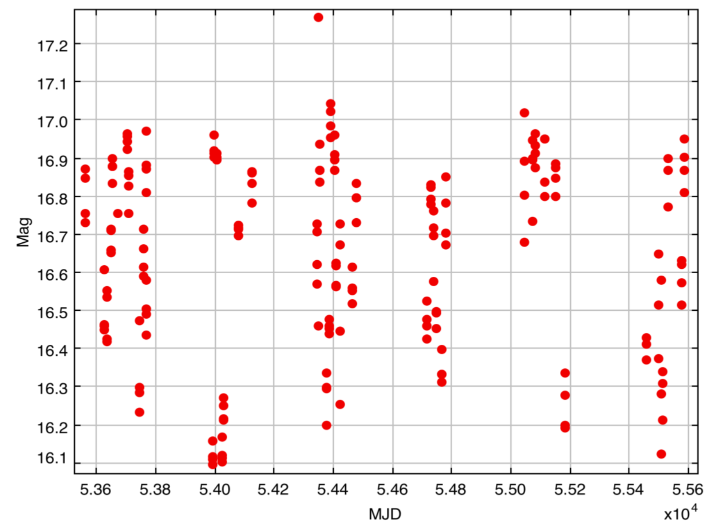
•Cluster measures

•Other

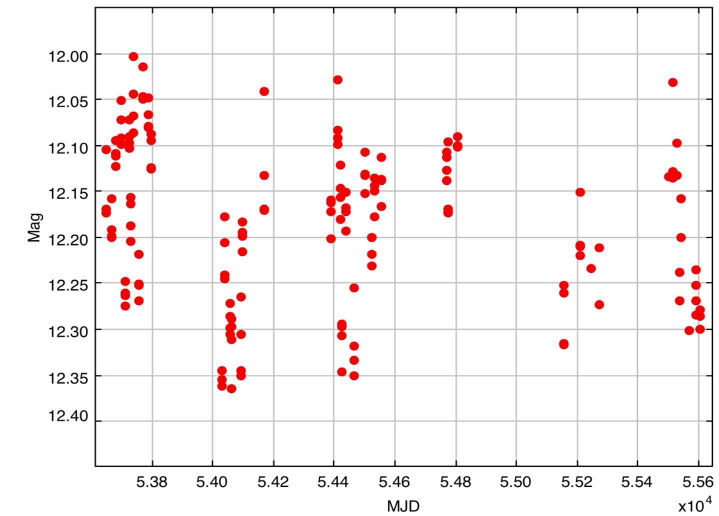
Over to Part 2 of Notebook

Features for RR Lyrae and W UMa

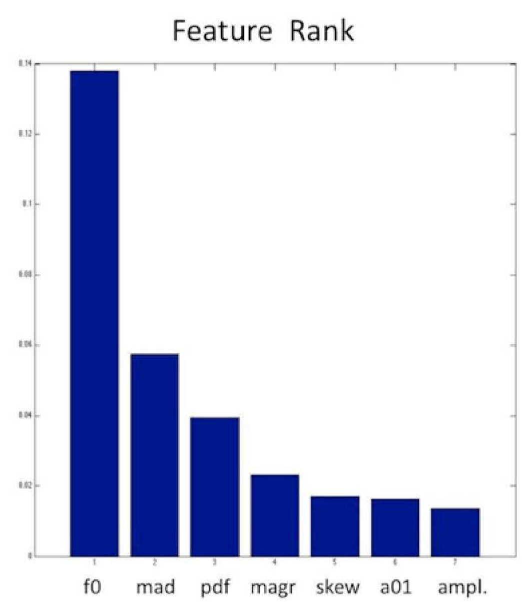
Rank features in the order
of classification quality for
a given classification problem,
e.g., RR Lyrae vs. WUMa

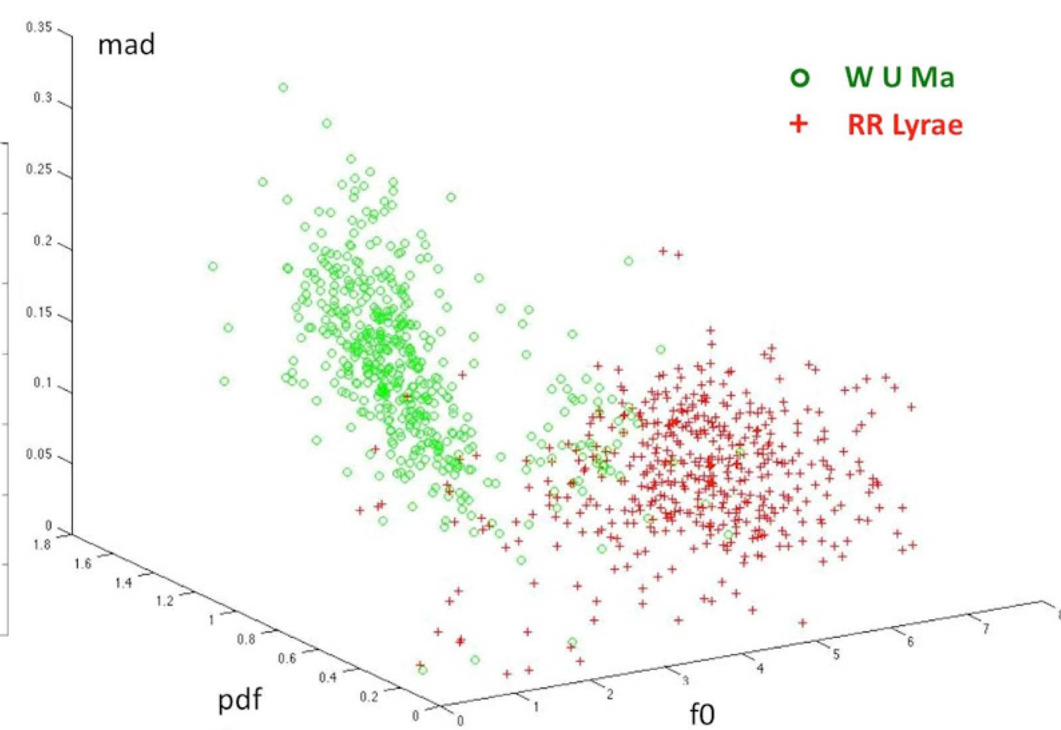


RR Lyrae



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

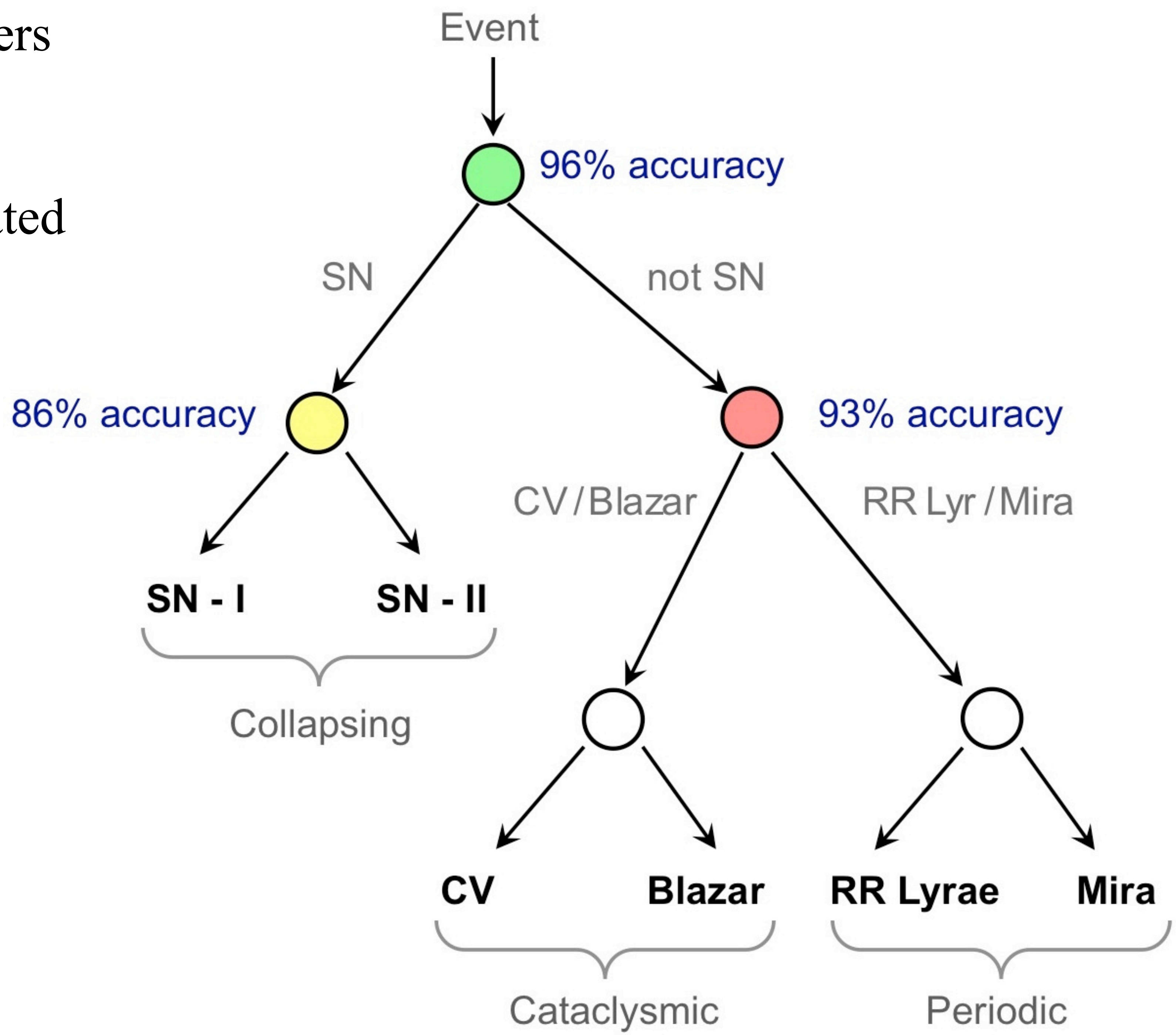
Fit the means within the groups (up to 4 measurements); and then take the logged SD of the residuals from this fit; the max absolute residuals from this fit; total variation of curve based on group means scaled by range of observation

A Hierarchical Approach to Classification

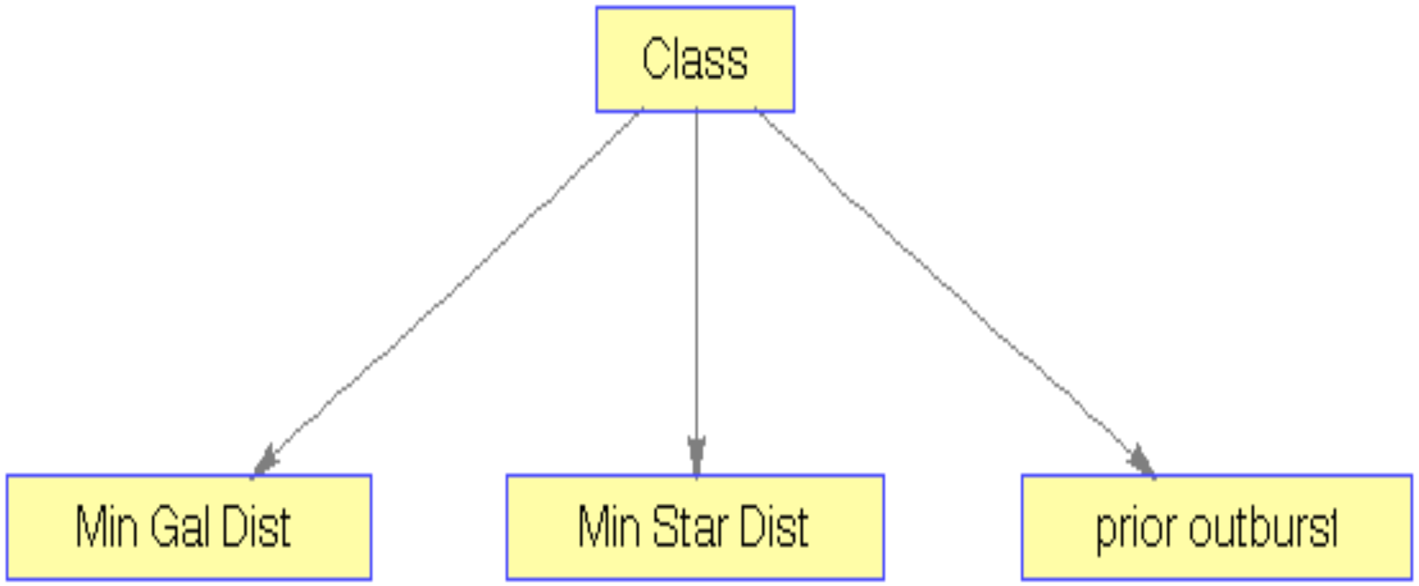
Different types of classifiers perform better for some event classes than for the others

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



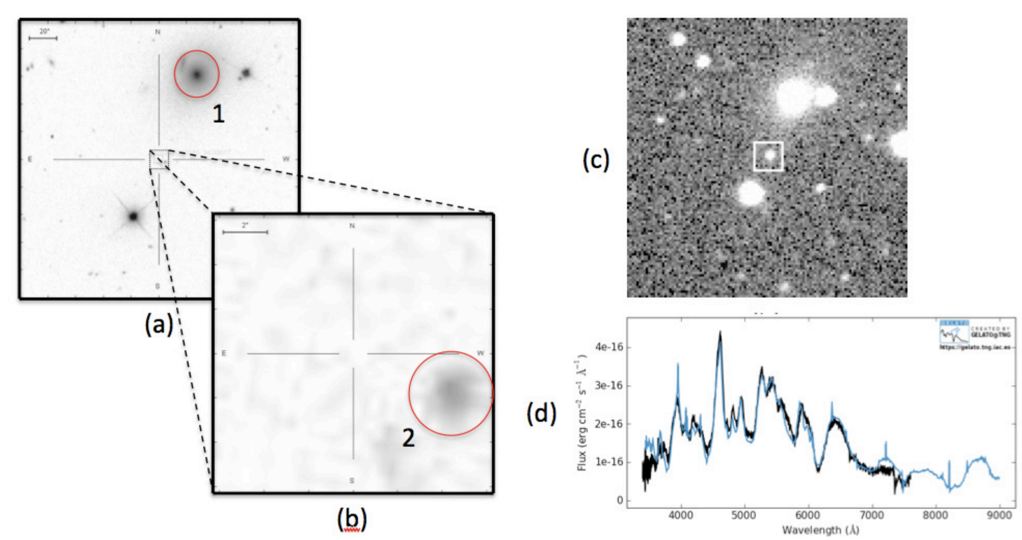
SN v. non-SN



normalized

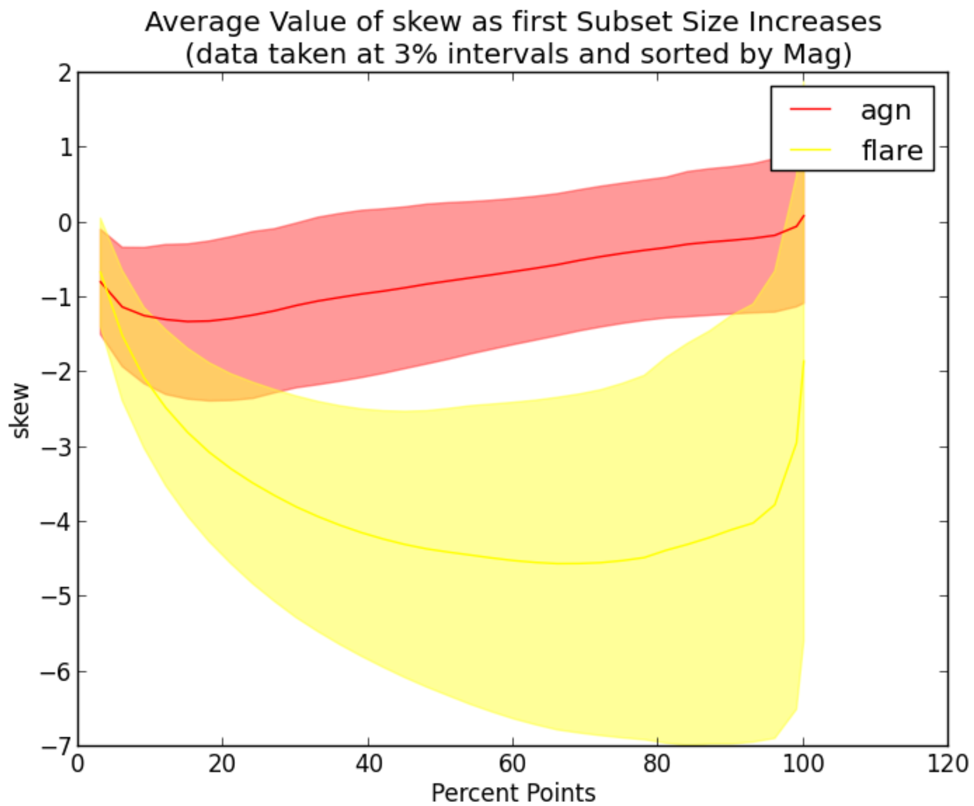
based on peaks

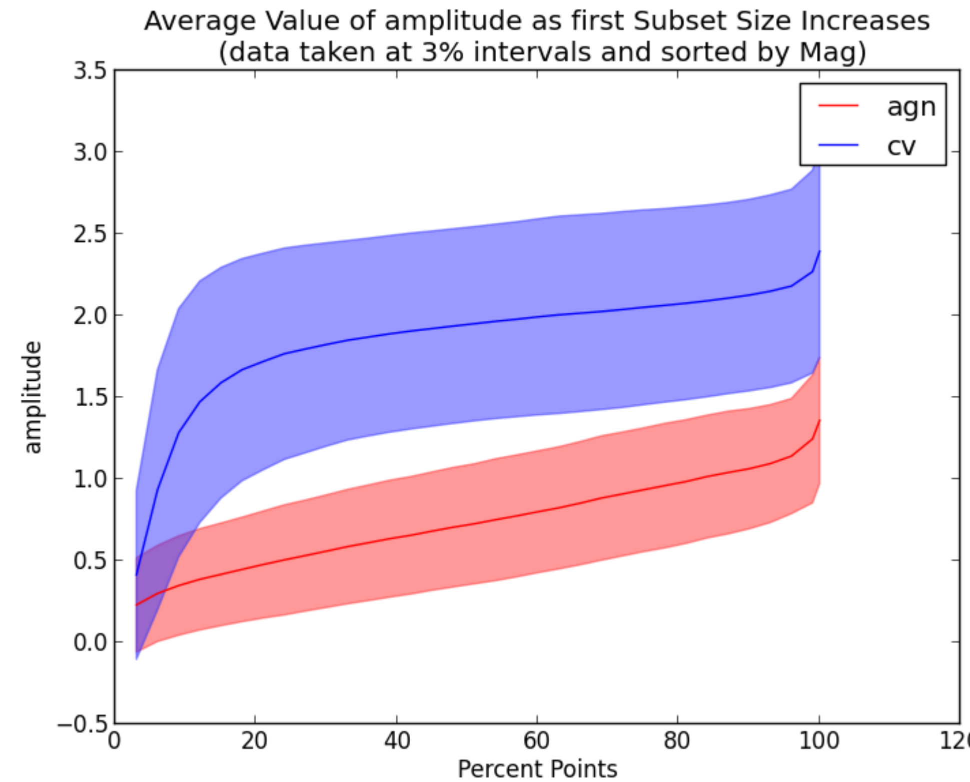
$$(\frac{1}{t_{span}}(\frac{1}{N}\Sigma_i w_i(p_i - p_m)^2)^{1/2}$$

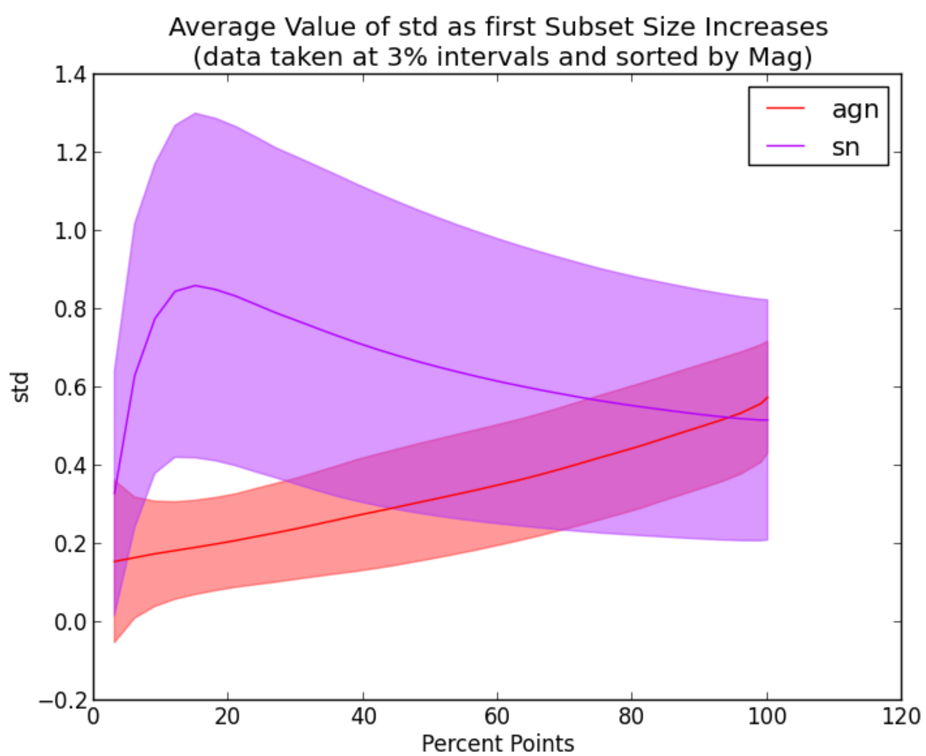


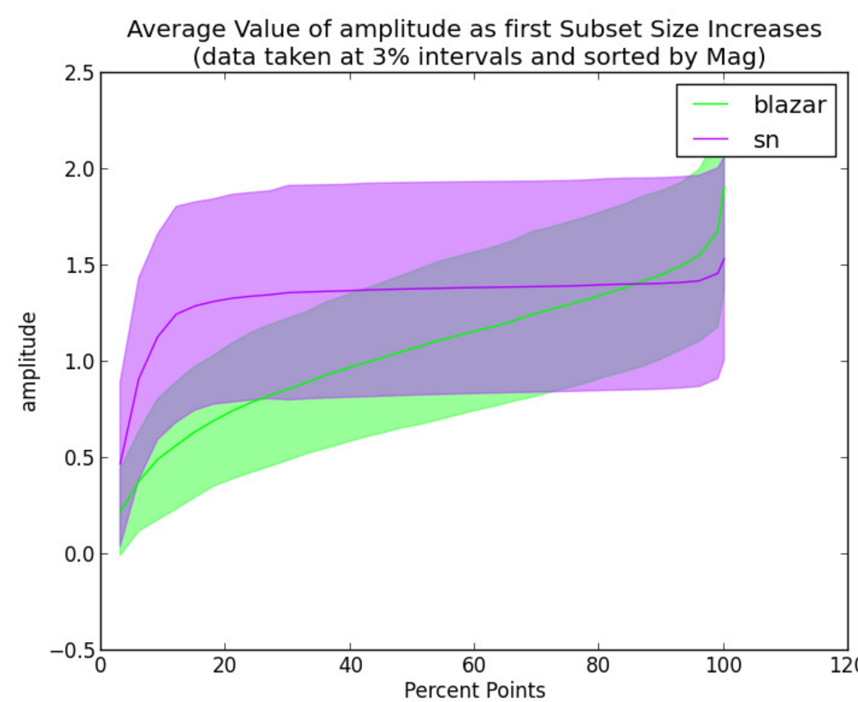
Using Discriminating Features for Brokering

Chengyi Lee



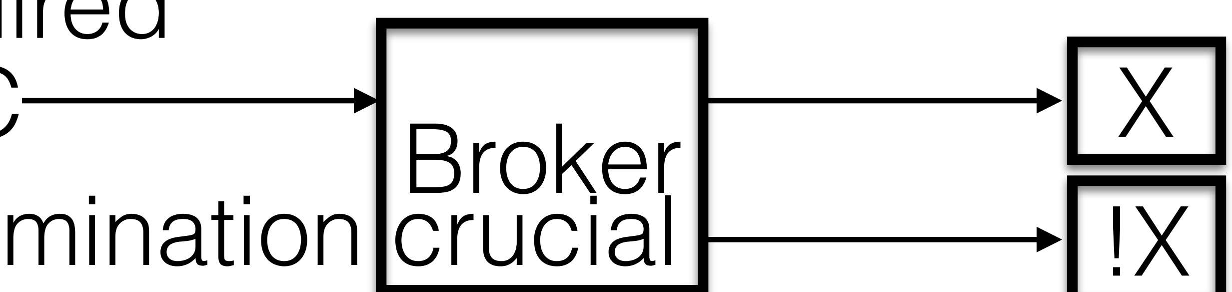






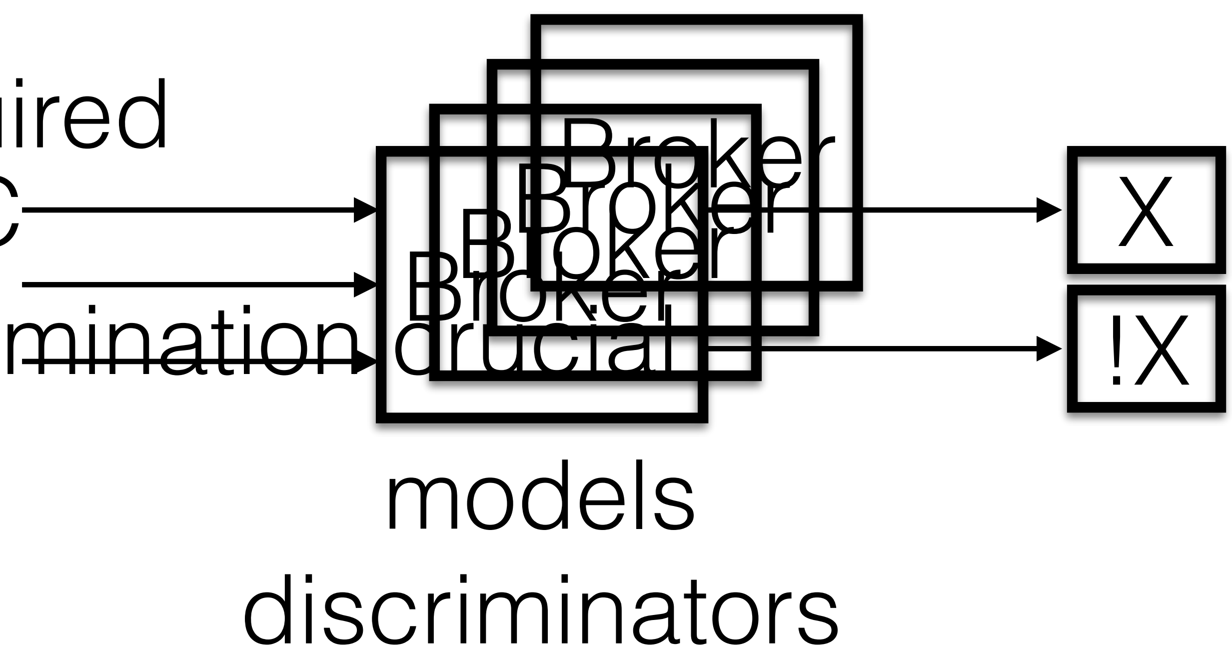
Binary Broker(s)

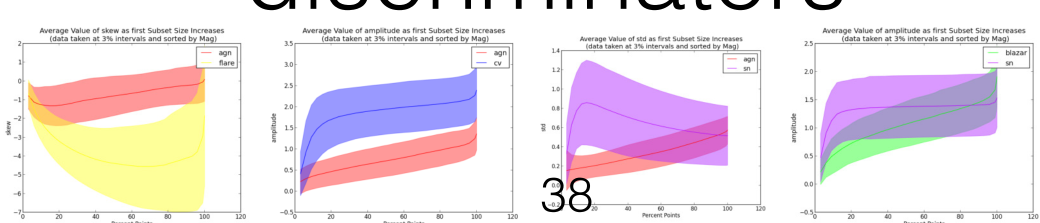
- Using features to tell classes apart - one class at a time
- Speed required
- Rarity determination crucial



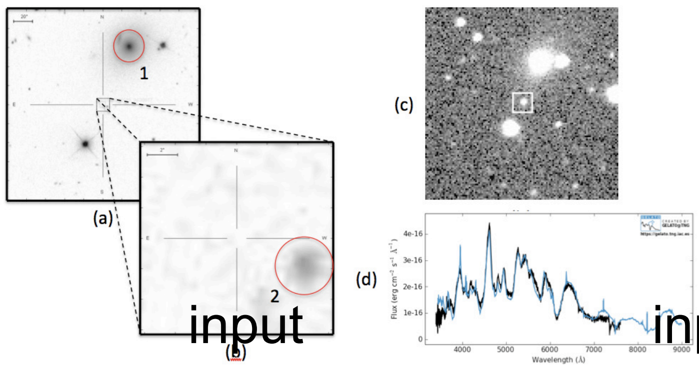
Binary Broker(s)

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





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

+

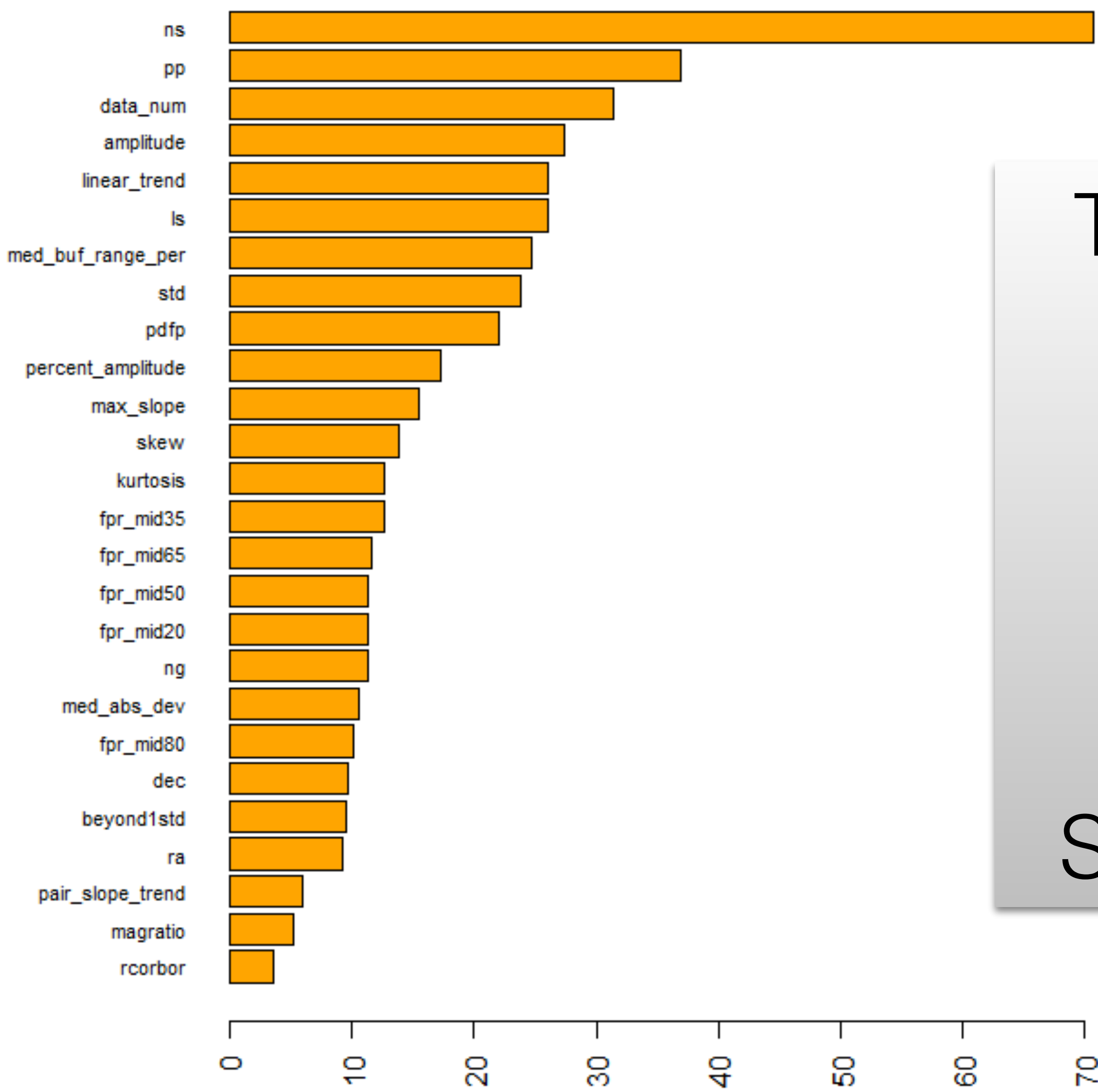
input — Not C4 / C4

input — Not C5 / C5

Extendible

Over to Part 3 of Notebook

Feature selection strategy



Take 2 variables at a time

Add 1 variable at a time

Start with all and reduce 1

Donalek et al. arxiv:1310.1976

Also PCA

t-distributed stochastic neighbor embedding (t-SNE)

van der Maaten, L.J.P.; Hinton, G.E. (2008)

$$p_{j|i} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|\mathbf{x}_i - \mathbf{x}_k\|^2 / 2\sigma_i^2)} \qquad p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N}$$

x_i, x_j: highdim objs
p_ij: similarity measure

$$q_{ij} = \frac{(1 + \|\mathbf{y}_i - \mathbf{y}_j\|^2)^{-1}}{\sum_{k \neq m} (1 + \|\mathbf{y}_k - \mathbf{y}_m\|^2)^{-1}}$$

Q: lower dimensional space

$$KL(P||Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

Minimize divergence between P and Q

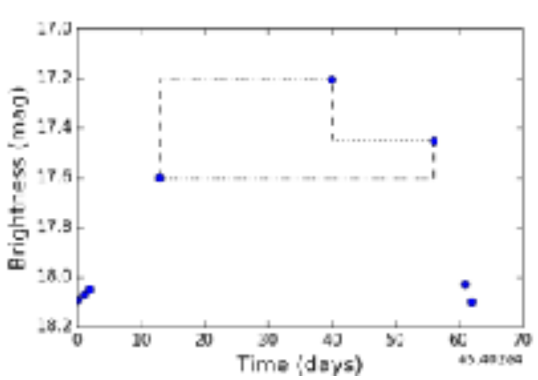
Over to Part 4 of Notebook

Many challenges

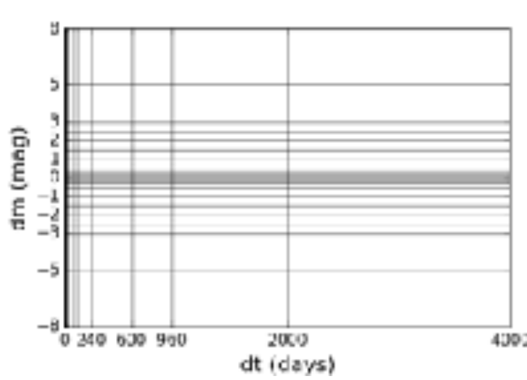
1. **Characterize/Classify as much with as little data as possible**
2. **Only a small fraction are rare - find/characterize them early**
3. **A variety of parameters - choose judiciously**
4. **Real-time computation is required - find ways to make that happen**
5. **Metaclassification - combining diverse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

dmdt, deep learning etc.

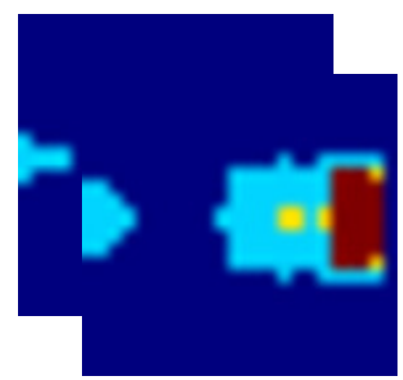
I deliberately did not go in to the more advanced topics related to image representation of light curves and using deep learning. Feel free to take that up as an exercise.



Light curves



Density representation

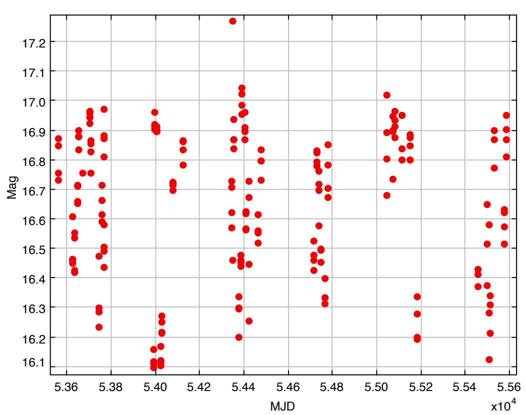


Equi-area images

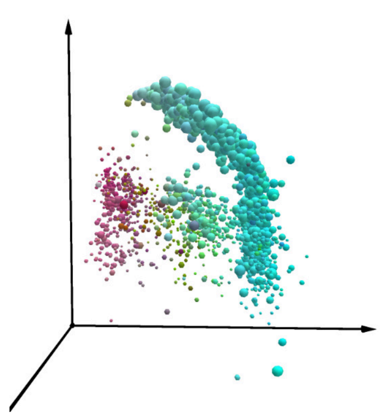
Convolutional Neural Network

Summary

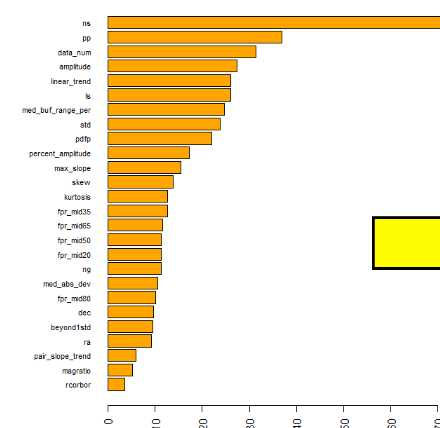
- Light curves -> many features
- Visualization/computation/choice
- Many features -> fewer features
- Classification



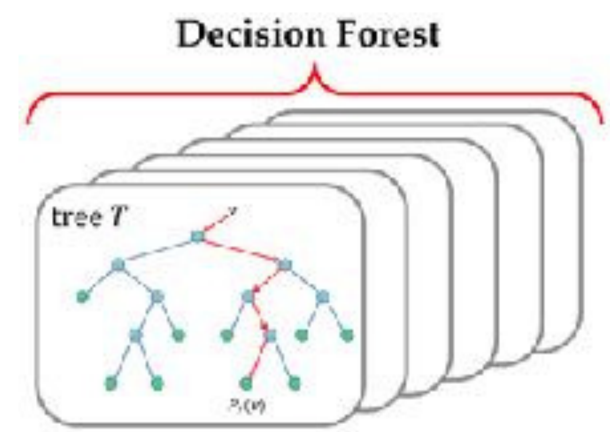
Light curves



Feature vectors



Dimensionality Reduction



Classification

A few References

- Cody: https://arxiv.org/abs/1401.6582
- Drake: https://arxiv.org/abs/1405.4290
- Faraway: https://arxiv.org/abs/1401.3211
- Graham: https://arxiv.org/abs/1306.6664
- Mahabal: https://arxiv.org/abs/0802.3199
- Richards:https://arxiv.org/abs/1101.1959
- **Many many others**