Automated classification of transients

Ashish Mahabal
aam@astro.caltech.edu
7 June 2011
KISS, Digging Deeper workshop tutorial

Collaborators

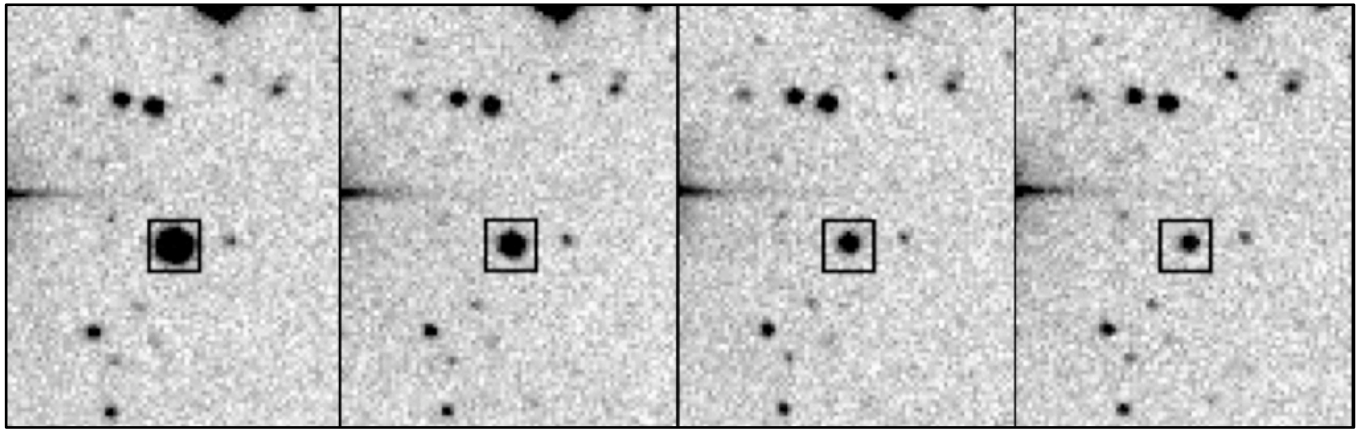
- Caltech
  - **George Djorgovski**
  - Ciro Donalek
  - Andrew Drake
  - Matthew Graham
  - Roy Williams
  - Nihar Sharma
  - Yutong Chen
- JPL
  - Baback Moghaddam
  - Mike Turmon

Plus at various other institutes all over, but especially in US, India and Italy

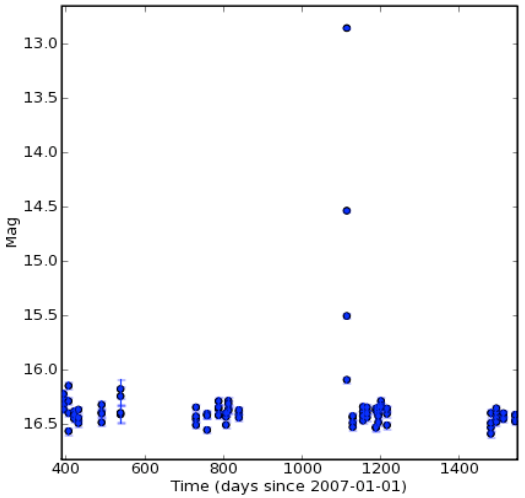
http://pardington10.wikis.birmingham.k12.mi.us/Collaboration+Techniques

What is a transient?

Example: 4 individual exposures, separated by 10 min



Light curve

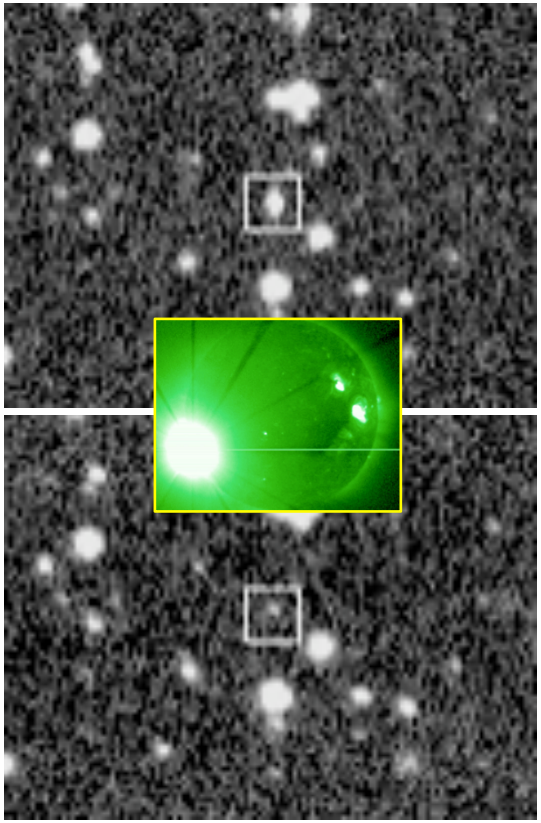


Fast transient (flaring dM), CSS080118:112149–131310

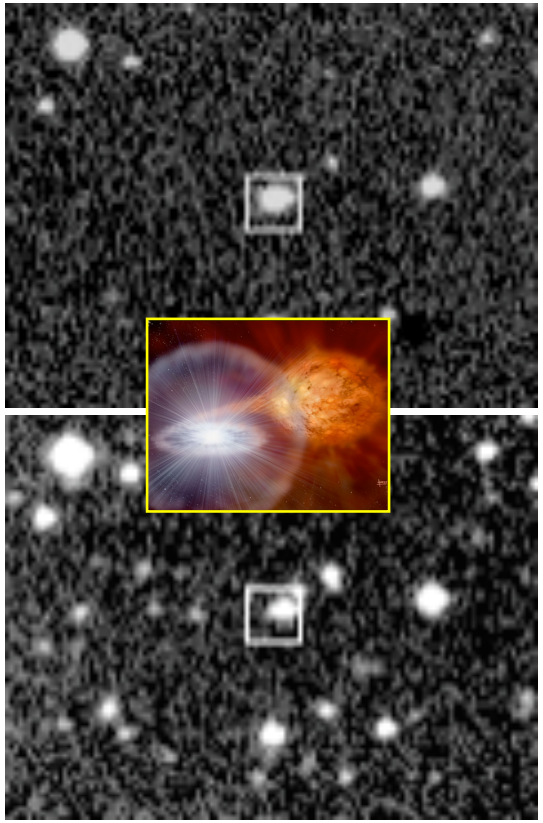
Something that has a large delta-magnitude for a small delta-time

Examples of CRTS Transients

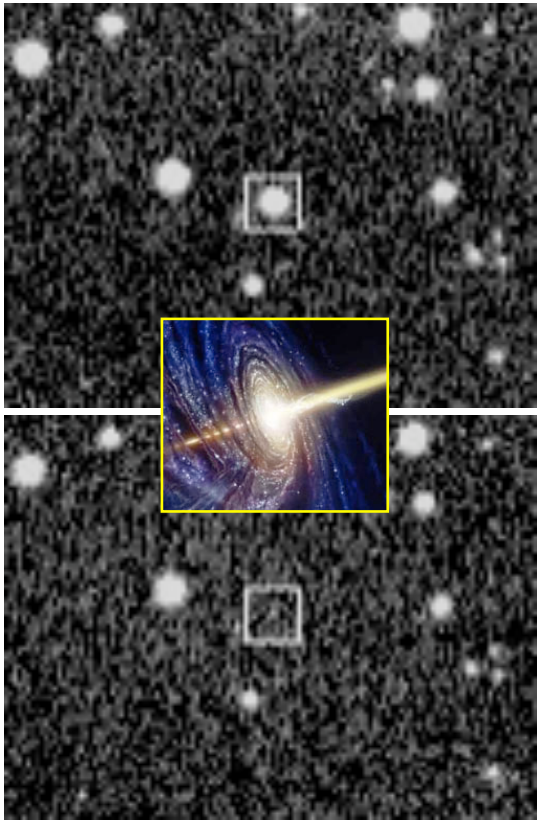
CSS090429:135125-075714
Flare star



CSS090429:101546+033311
Dwarf Nova



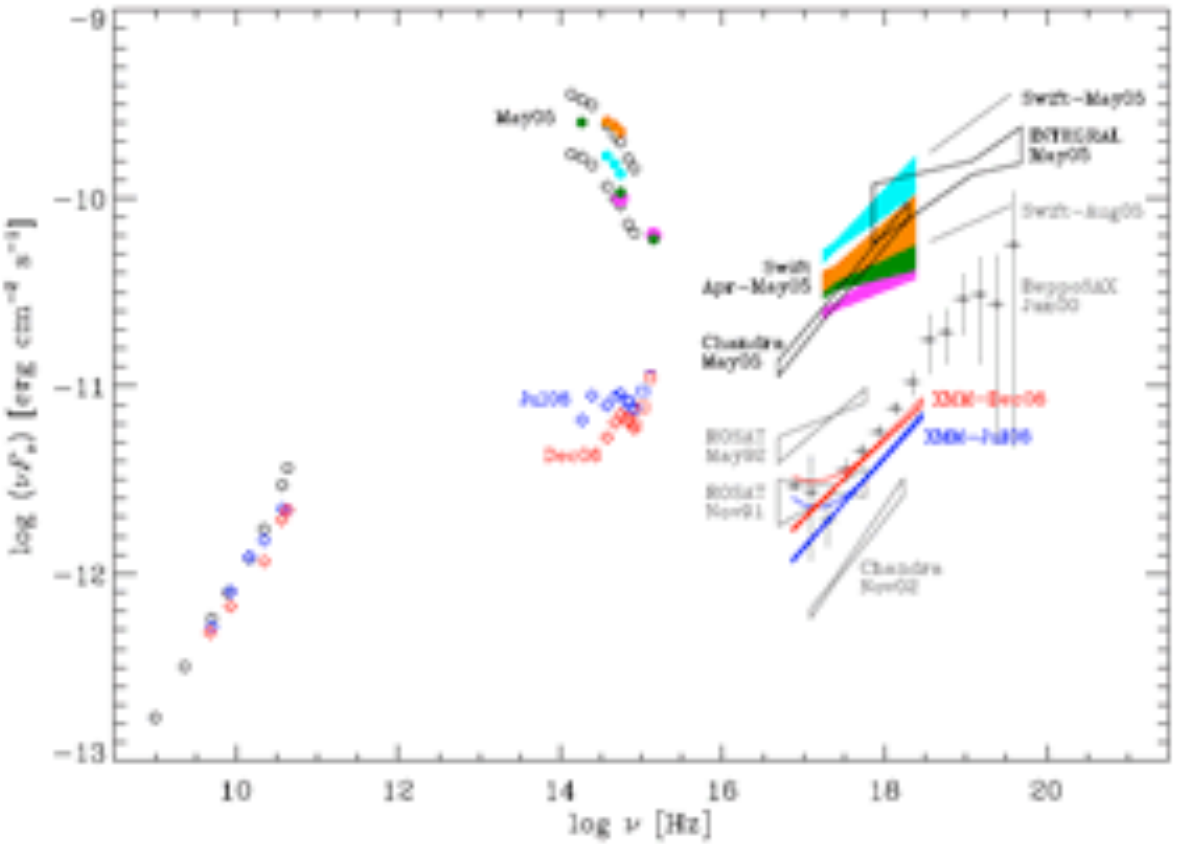
CSS090426:074240+544425
Blazar, 2EG J0744+5438



Vastly different physical phenomena, and yet they look the same!
Which ones are the most interesting and worthy of follow-up?

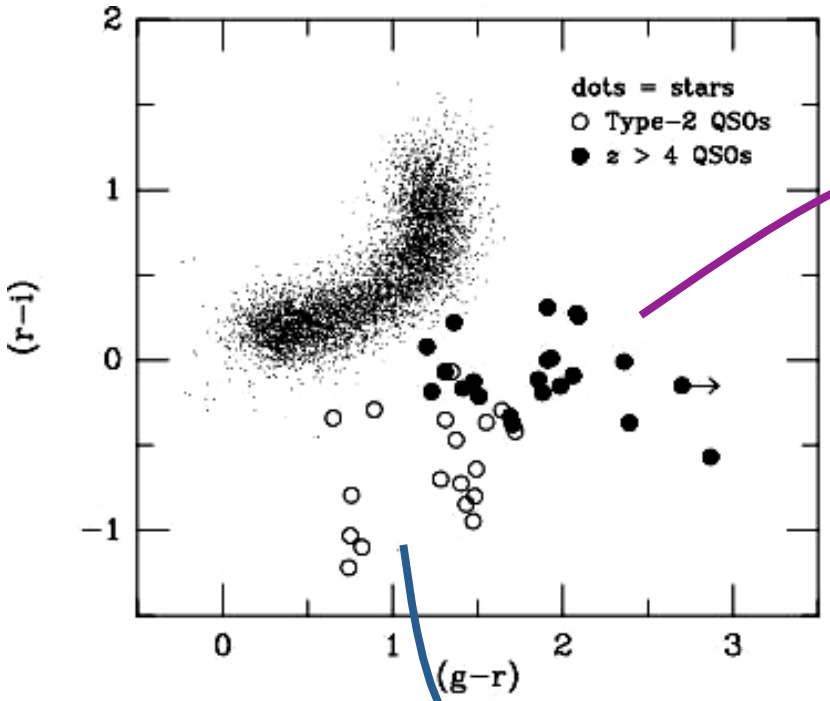
Rapid, automated transient classification is a critical need!

SED: Spectral Energy Distribution

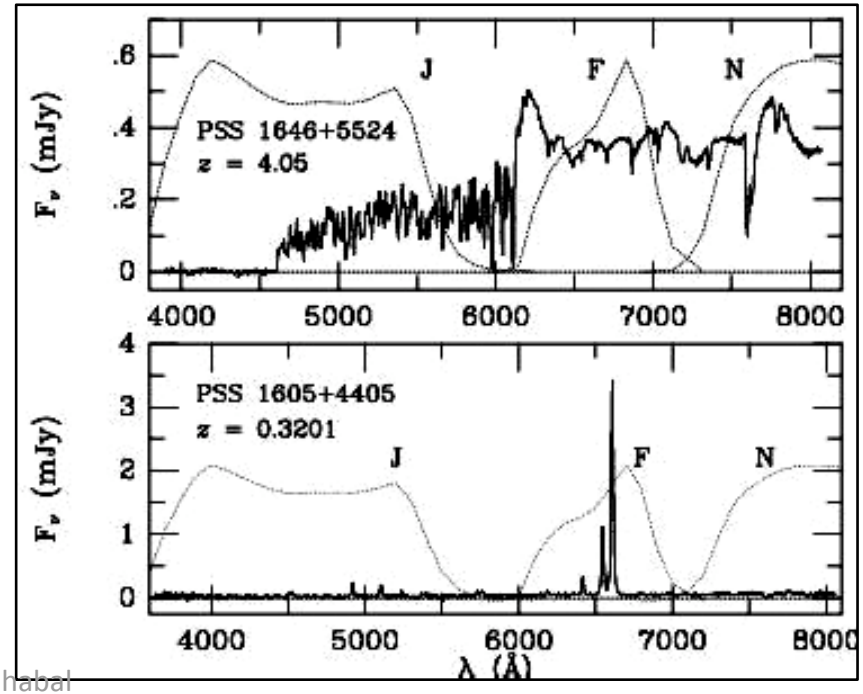


http://hera.ph1.uni-koeln.de/~heintzma/k1/3C273_HST.htm

Colors (e.g. from follow-up from Palomar 60-inch) are a narrower aspect of the SED

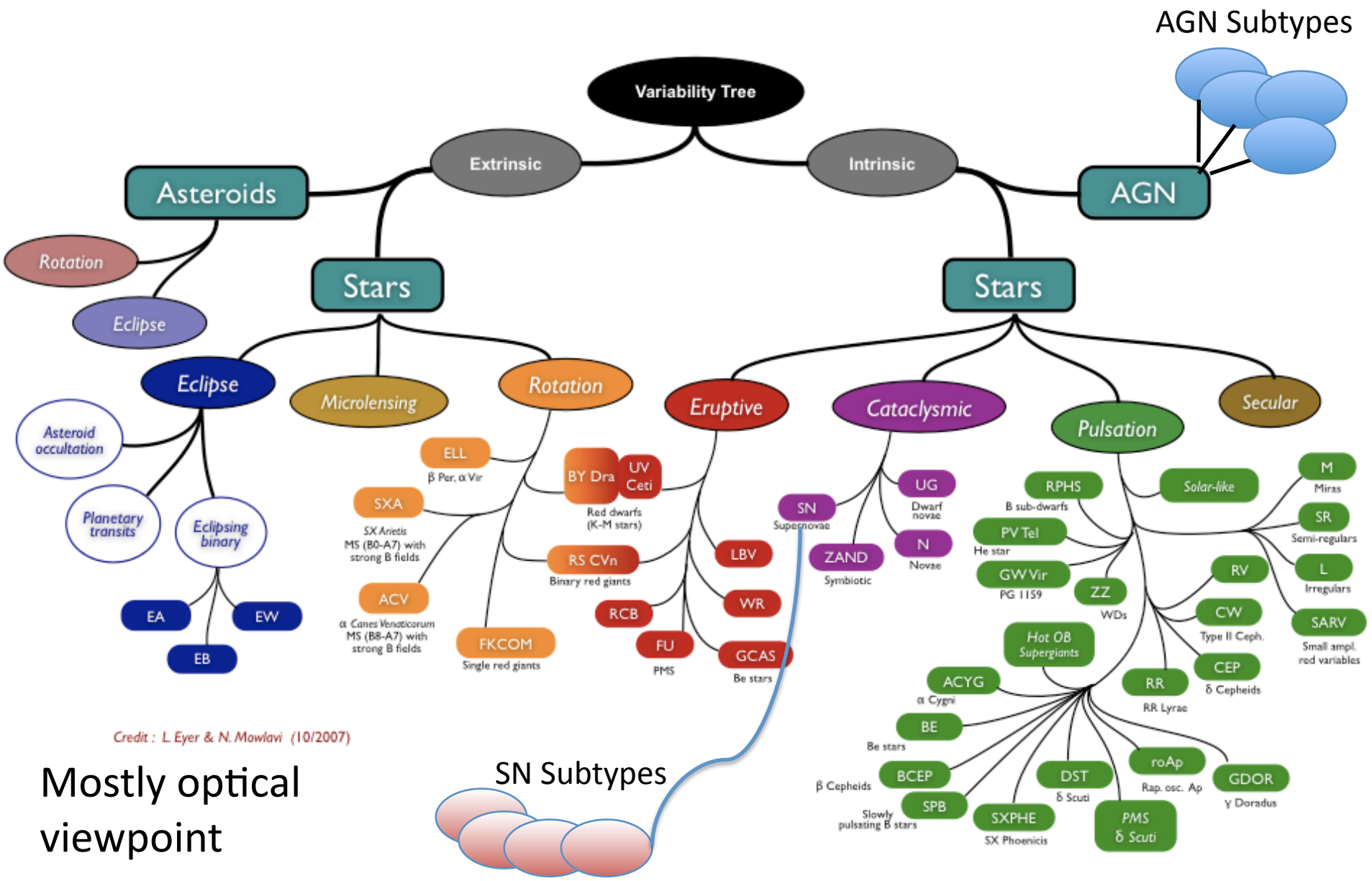


High-z QSO



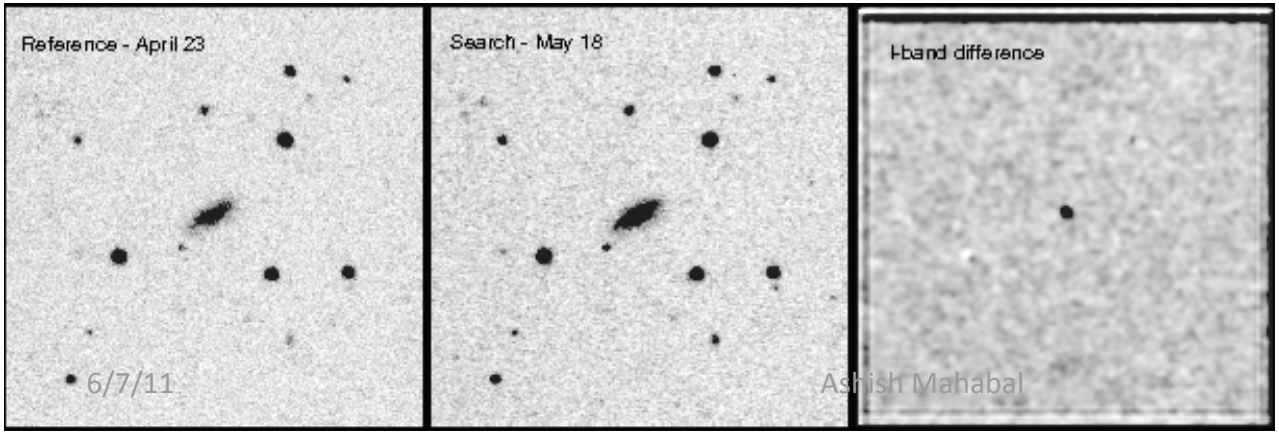
Type-2 QSO

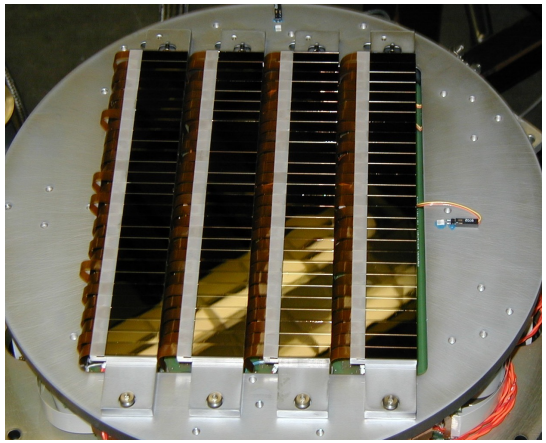
Semantic Tree of Astronomical Variables and Transients



The Palomar-Quest (PQ)
Digital Synoptic Sky Survey

- Palomar 48-in. + 112-CCD, 161 Mpix camera
- A Caltech-Yale collab. Co-PIs: C. Baltay & SGD; plus other groups worldwide (LBL, etc.)
- Many passes with up to 4 filters (*UBRI/griz*), time baselines from minutes to years
- Collected > 50 TB of data
- Operated from Aug. 2003 through Sept. 2008
- ***Key goal: Exploration of the time domain***





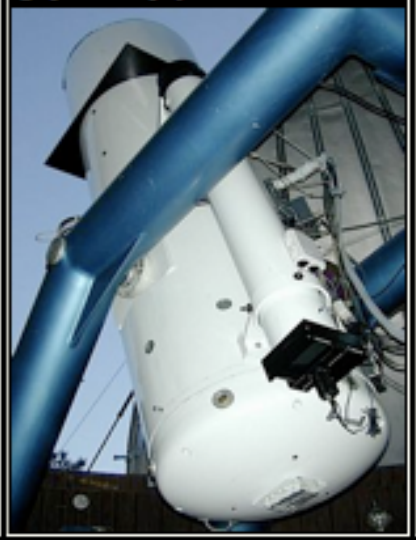
⇦ LBL SNF search (Nugent et al.)

> 700 SNe discovered

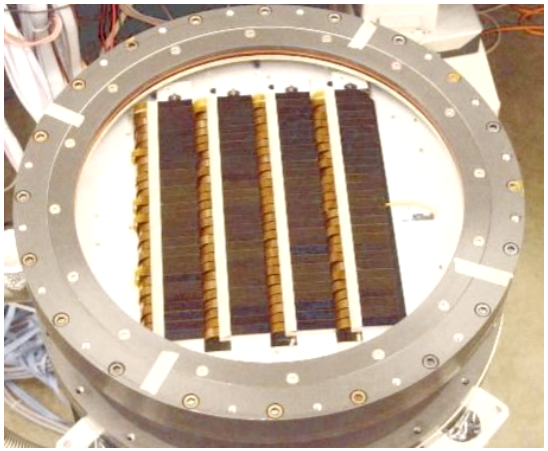
Catalina Sky Survey(s):

NEO survey Co-PI's:
E. Beshore & S. Larson (LPL)

CRTS uses the data from all three Catalina NEO surveys, with a coverage of up to 2,500 deg^2 / night, and the total area coverage of ~ 30,000 deg^2

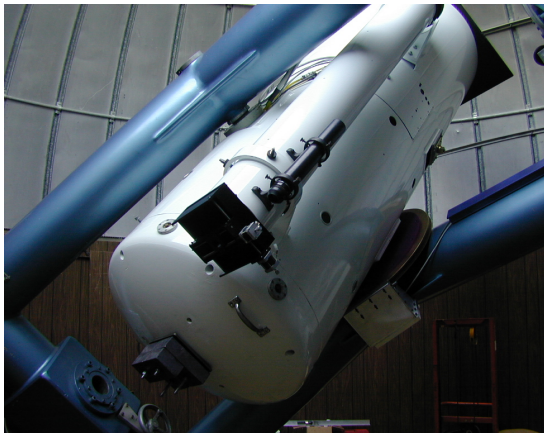
	MLS The Mt. Lemmon Survey 1.5m Cass	CSS Catalina Sky Survey 0.7m Schmidt	SSS Siding Springs Survey 0.5m Schmidt
			
Survey region (deg)	+/- 5 deg ecliptic	-25 < Dec < +70	-80 < Dec < -25
Field of View (square deg)	1.2	8.1	4.2
Mag limit (V)	21.5	19.5	19.0

We are processing the Catalina data streams in real time to look for astrophysical transients



PQ

CSS



GALEX, Spitzer, FIRST, …

Recent, current and future multiepoch surveys

Hundreds of thousands of transients per night in the near future

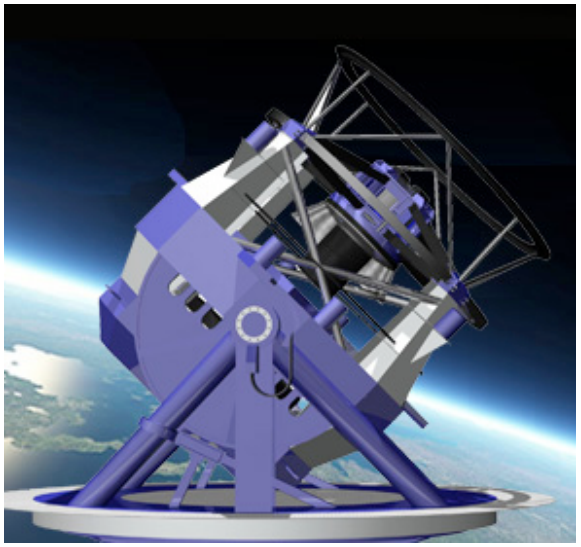
PTF GW

Skymapper

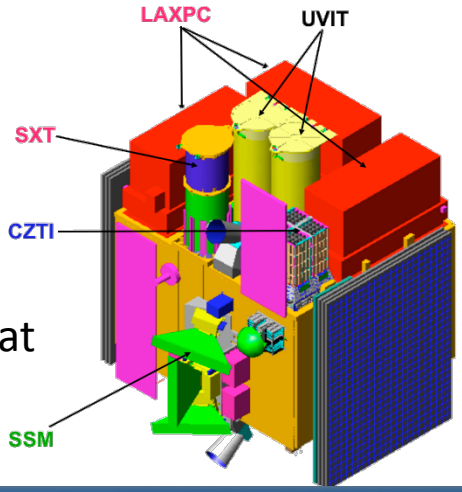
Pan-STARRS

Orders of magnitudes different.

Move towards digital movies!



LSST



Astrosat

CRTS Event Detections

A Drake

Distinct Events Detection Statistics as of 5 Jun 2011 UT:

Tel	All OTs	SNe	CVs	Blazars	Ast/ flares	CV/ SN	AGN	Other
CSS	2033	596	501	113	184	275	229	195
MLS	1560	183	38	12	122	374	744	214
SSS	227	24	93	7	5	43	16	42
Total	3820	803	632	132	311	692	989	451

- Threshold set deliberately very high – only the most dramatic transients are pulled out in the real time
- About 1 strong transient per 10^6 source detections
- The rate of significant transients/variables is at least an order of magnitude higher
- Many events are re-detected repeatedly (not counted above)

The Palomar-Quest Event Factory

Sept. 2006

Detect ~ 1 - 2 $\times 10^6$ sources per half-night scan

Compare with the baseline sky

Find ~ 10^3 apparent transients (in the data)

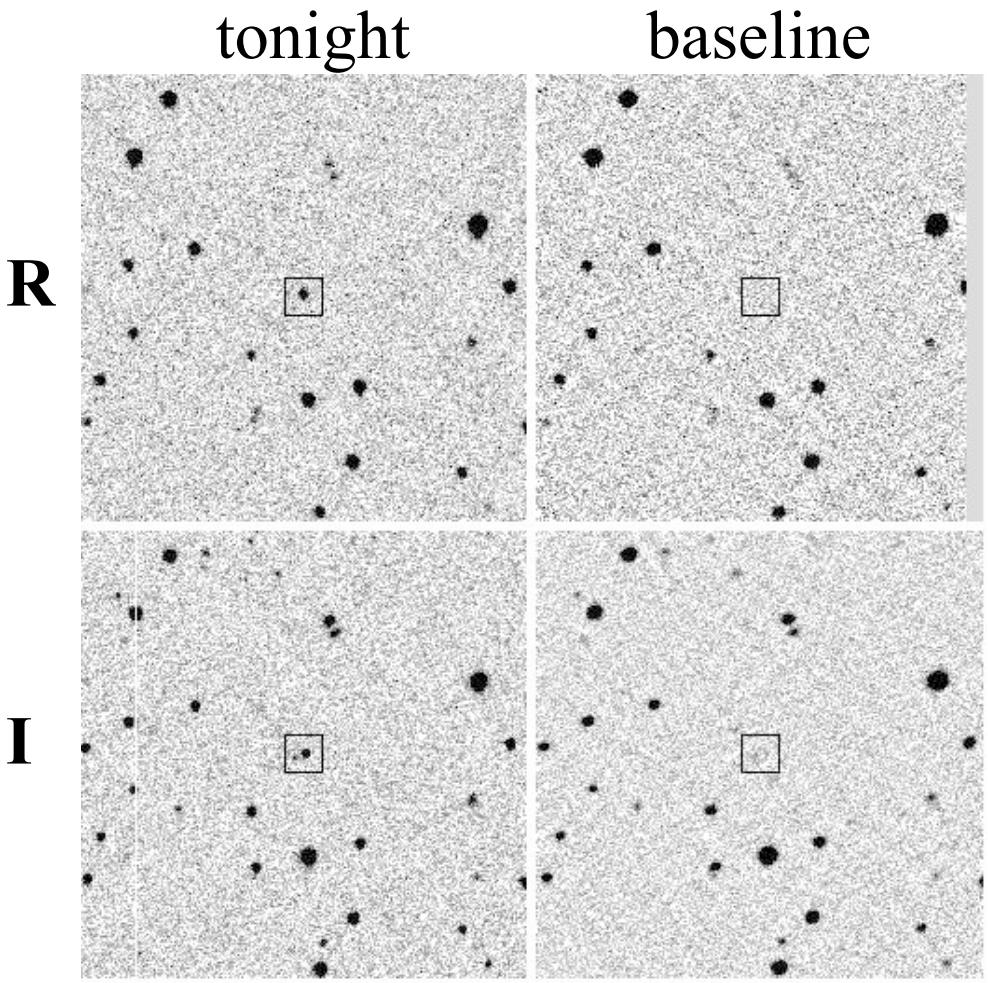
Remove instrum. artifacts

Identify ~ 2 - 4 $\times 10^2$ real transients (on the sky)

Remove asteroids

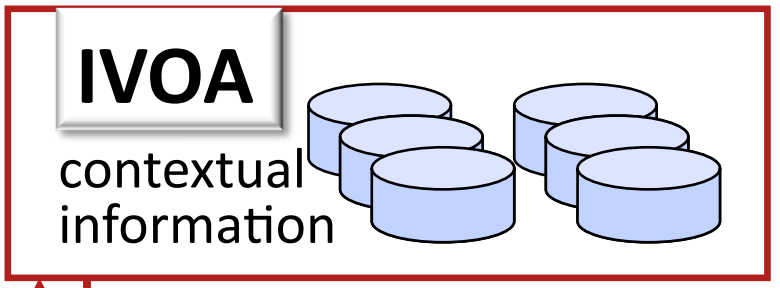
Identify ~ 1 - 10 possible Astrophysical transients

Classification and follow-up



Towards Automated Event Classification

A **necessity** for large synoptic surveys



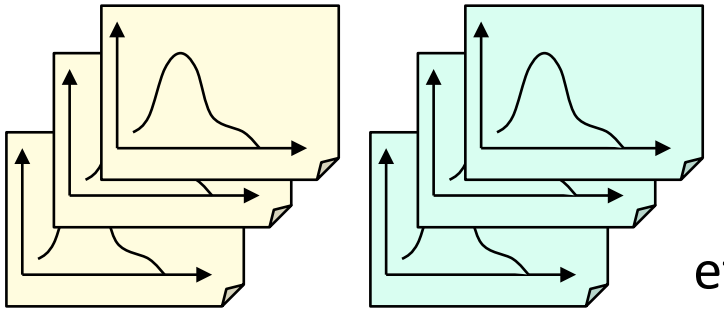
Event parameters:
$m_1(t)$, $m_2(t)$,…
α, δ, μ, …
image shape…

Event Classification Engine

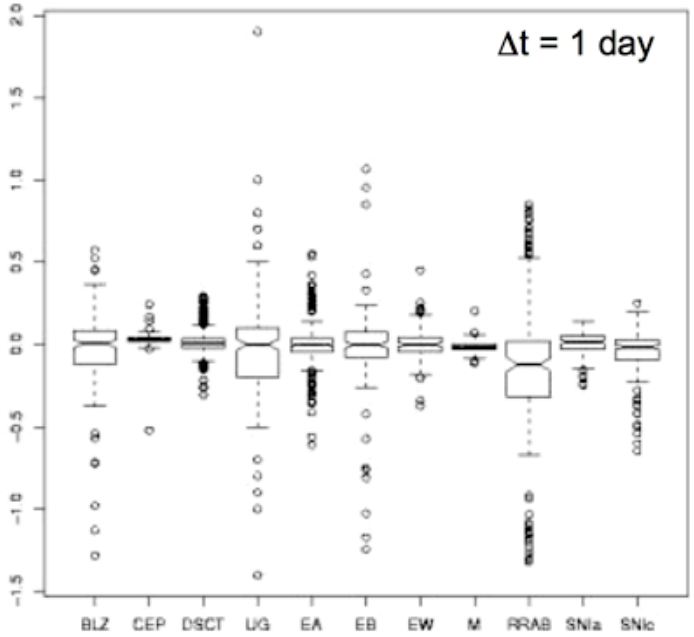
P(SN Ia) = …
P(SN II) = …
P(AGN) = …
P(CV) = …
P(dM) = …
….

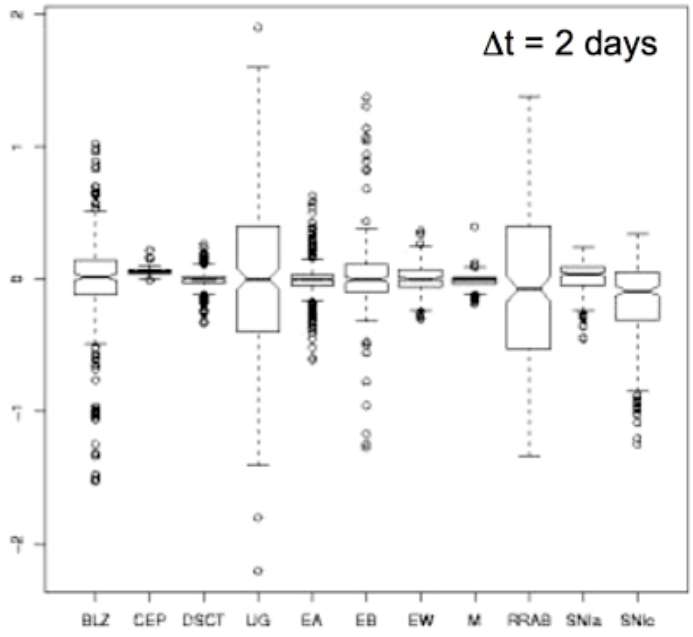
Classification probabilities (evolving, iterated)

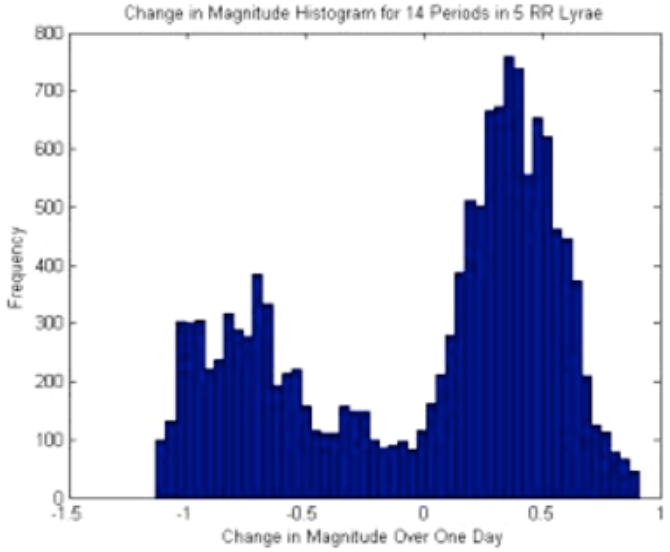
Expert and ML generated priors

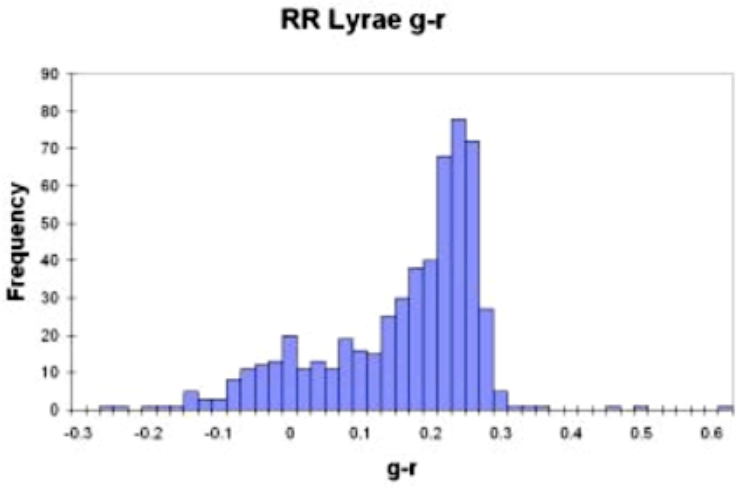


With M Turmon and B Moghaddam, JPL

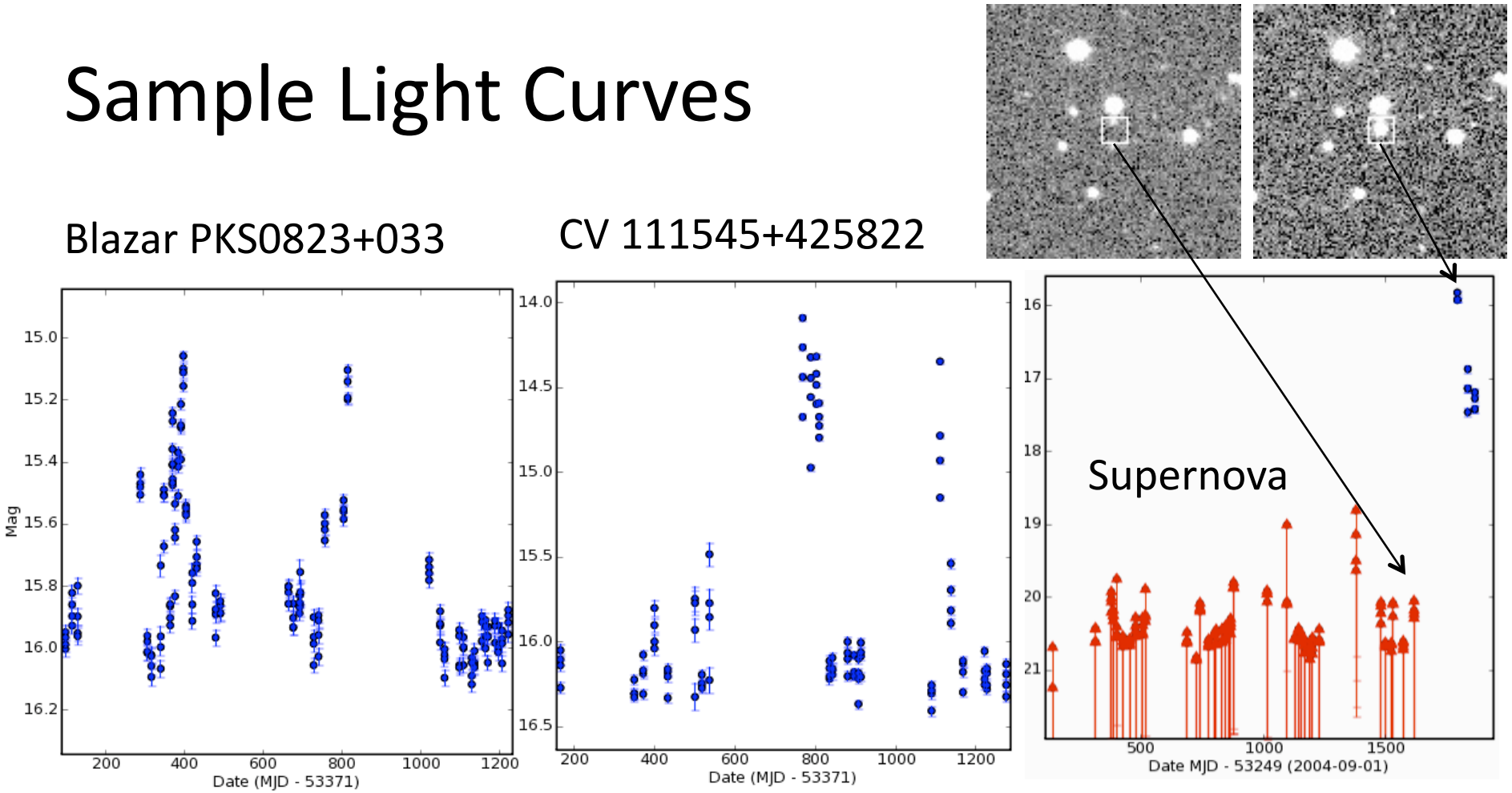

Δt = 1 day
BLZ
CEP
DSCT
UG
EA
EB
EW
M
RRAB
SNIa
SNIc


Δt = 2 days
BLZ
CEP
DSCT
UG
EA
EB
EW
M
RRAB
SNIa
SNIc


Change in Magnitude Histogram for 14 Periods in 5 RR Lyrae
Frequency
Change in Magnitude Over One Day


RR Lyrae g-r
Frequency
g-r

Sample Light Curves

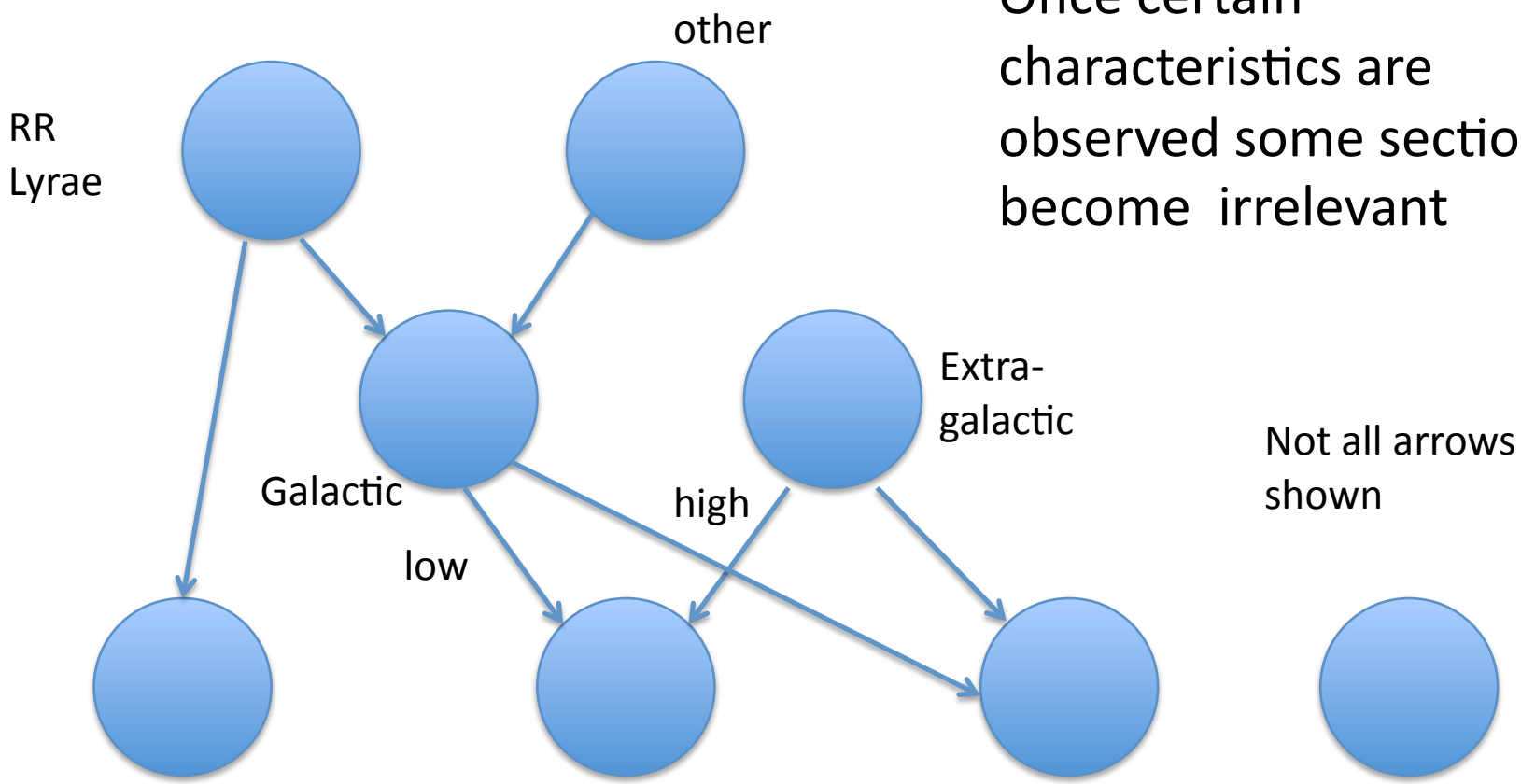


Variables and transients – the distinction is one of perception, and your aims

Building Bayesian Networks

- Handling of incomplete data
  - Real-world cases
- Learning causal connections
  - What variable caused what
- Incorporating domain knowledge
  - Experts can weight in at different points
- Memorizing (aka overfitting) avoided
  - No holdout necessary

BN is an Acyclic Directed Graph



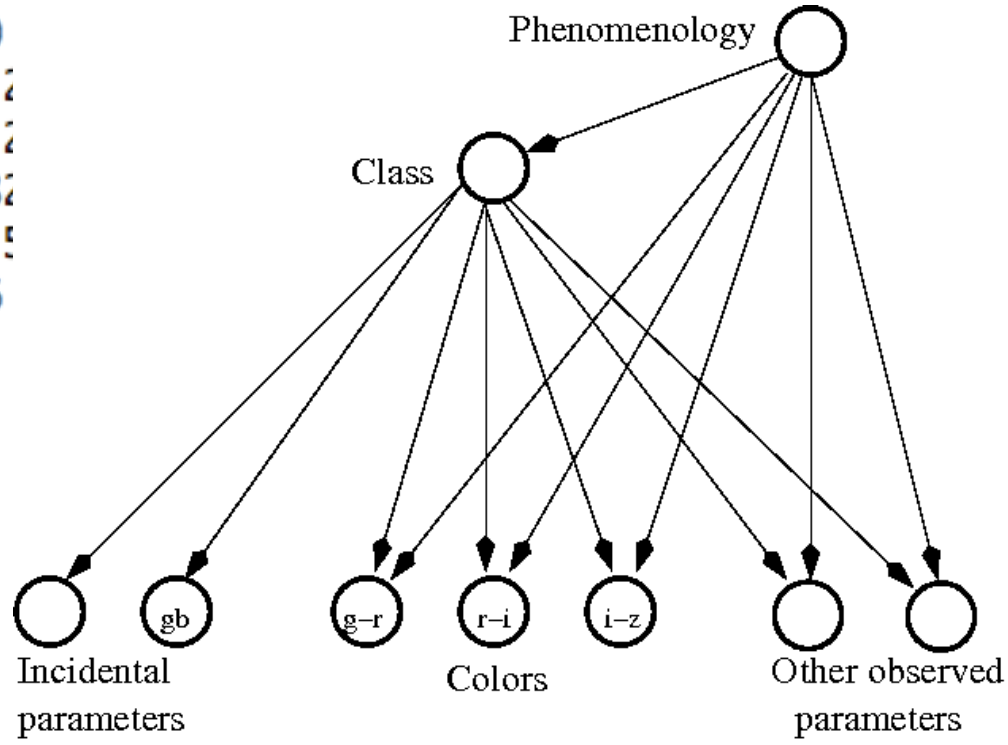
Once certain characteristics are observed some sections become irrelevant

Not all arrows shown

Sample data input to BN

C Donalek,
N Sharma

id	gminr	rmini	iminz	gb	class		pCV	pSN	pblazar
1	80130118012410358	6 0.20	0.49	-1.06	41.570266	1	0.433000	0.221294	0.343222
2	80130118012410358	6 0.72	0.43	0.30	41.570266	1	0.114421	0.130915	0.754664
3	80130123018414442	0 0.16	0.50	-0.30	25.068228	1	0.945996	0.015071	0.038933
4	80130123018414442	0 0.18	0.54	-0.38	25.068228	1	0.959667	0.024743	0.015591
5	80130123018414442	0 0.19	-99.0	-99.0					
6	80130123018414442	0 1.01	0.69	0.55	2				
7	80130123018414442	0 1.72	0.69	-0.07	2				
8	80201132055410799	6 -0.70	-0.16	-0.8	2				
9	80219123075411438	0 0.76	0.14	-0.02	5				
10	80219123075411438	0 0.79	0.12	-0.16					



The output is BN class which is fed to skyalert as an annotation to the original event

Radio, for instance

CV/SN classification ~80% with single epoch

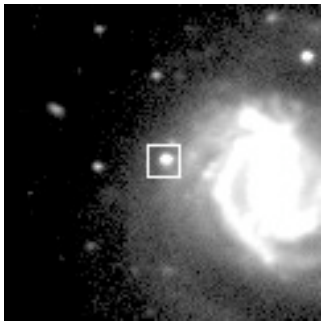
Naïve Bayes

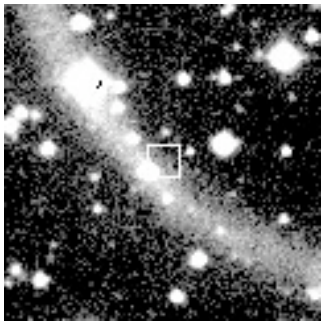
$$P(y = k \mid x) = P(x \mid y = k)P(k)/P(x) \propto P(k)P(x \mid y = k) \approx P(k)\prod\prod_{b=1}^{B} P(x_b \mid y = k)$$

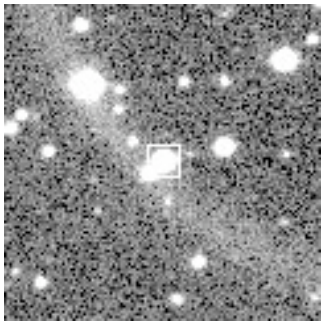
- x: feature vector of event parameters
- y: object class that gives rise to x (1<y<k)
- Certain features of x known: (position, flux)
- Others will be unknown: (color, delta-mag)
- Assumption: based on y, x is decomposable into B distinct independent classes (labeled x_b)
- This helps with the curse of dimensionality
- Also allows us to deal with missing values

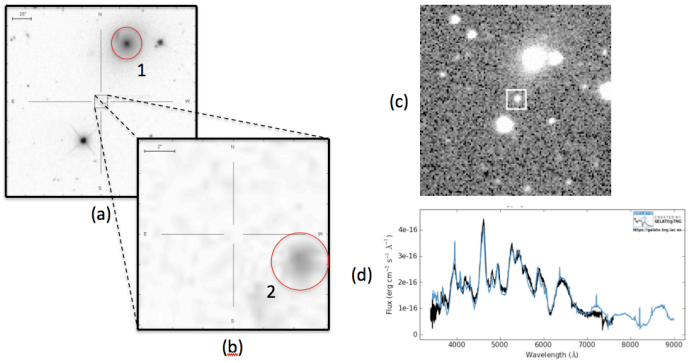
The importance of context

Which galaxy does a supernova belong to?

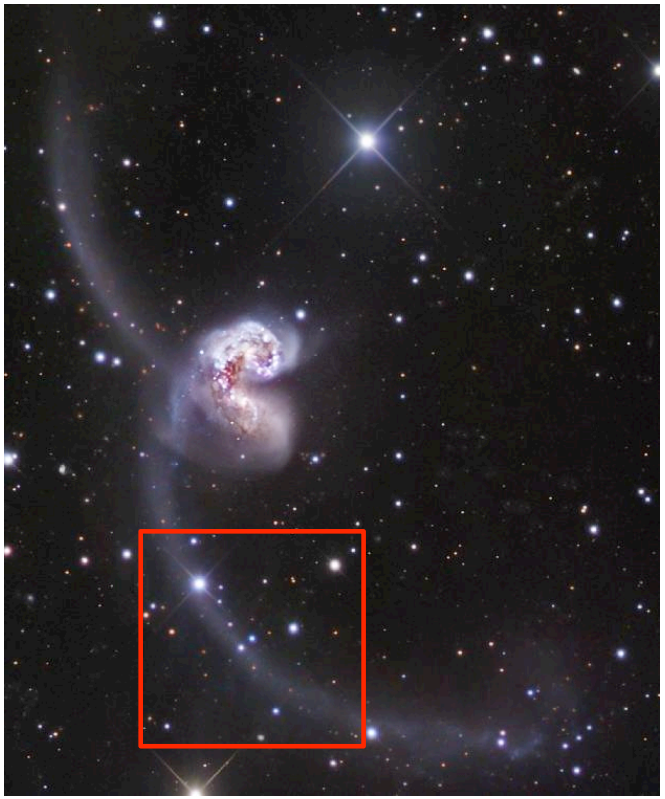






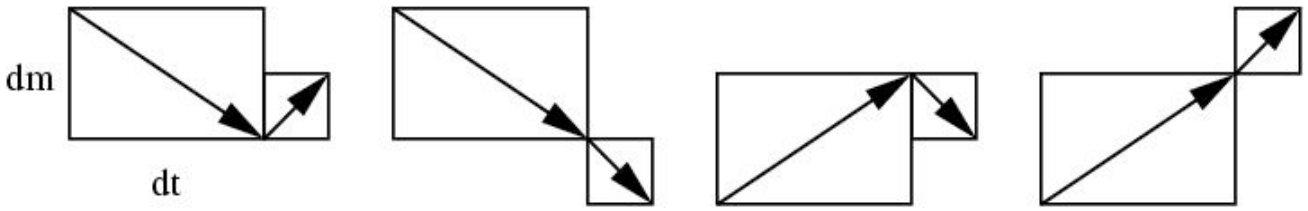


The need to see the big picture



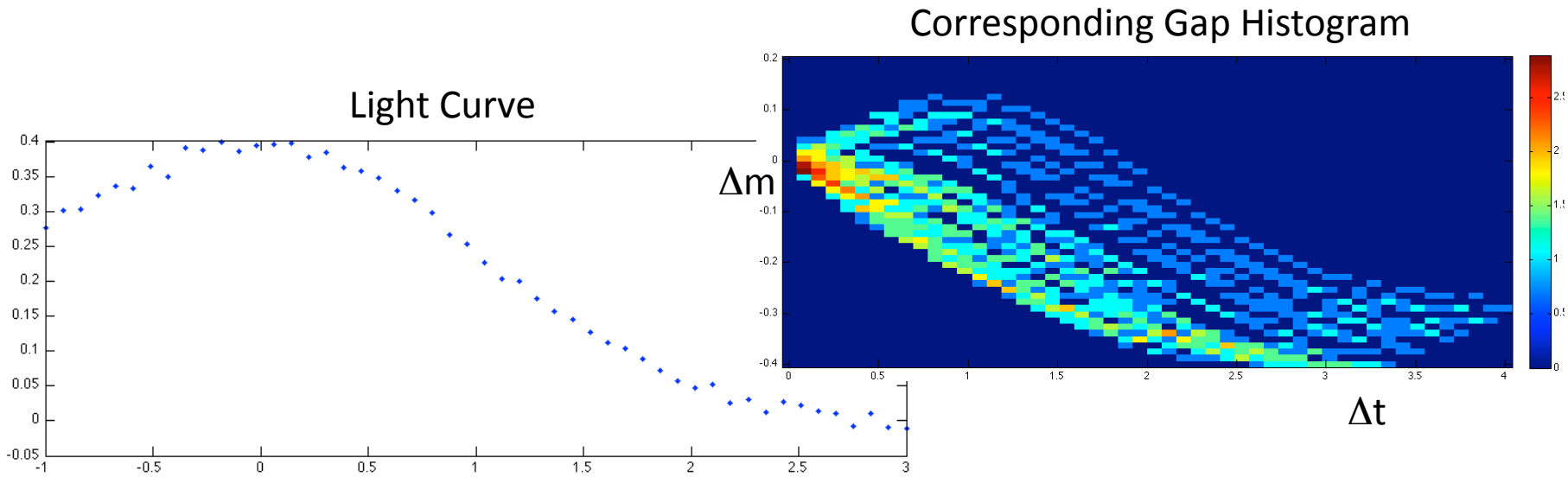
Characterization Vs. Classification

- Early focus on the extraction and dissemination of time series
- Characterizations is important
  - dm/dt
  - change of direction per unit time
  - change in periodicities (e.g,. wavelet or fourier decomposition);
  - variation in dm/dt
  - acceleration in dm/dt



Most SNe will not become fainter and then brighten up

Aspects of “Gap” processing



- Gap features capture sparse or irregular LCs
- The features, and thus the underlying density models, are invariant to absolute magnitude and time shifts
- Features & densities allow bound-only flux observations
  - Under poor seeing, we obtain only bounds like m > 18

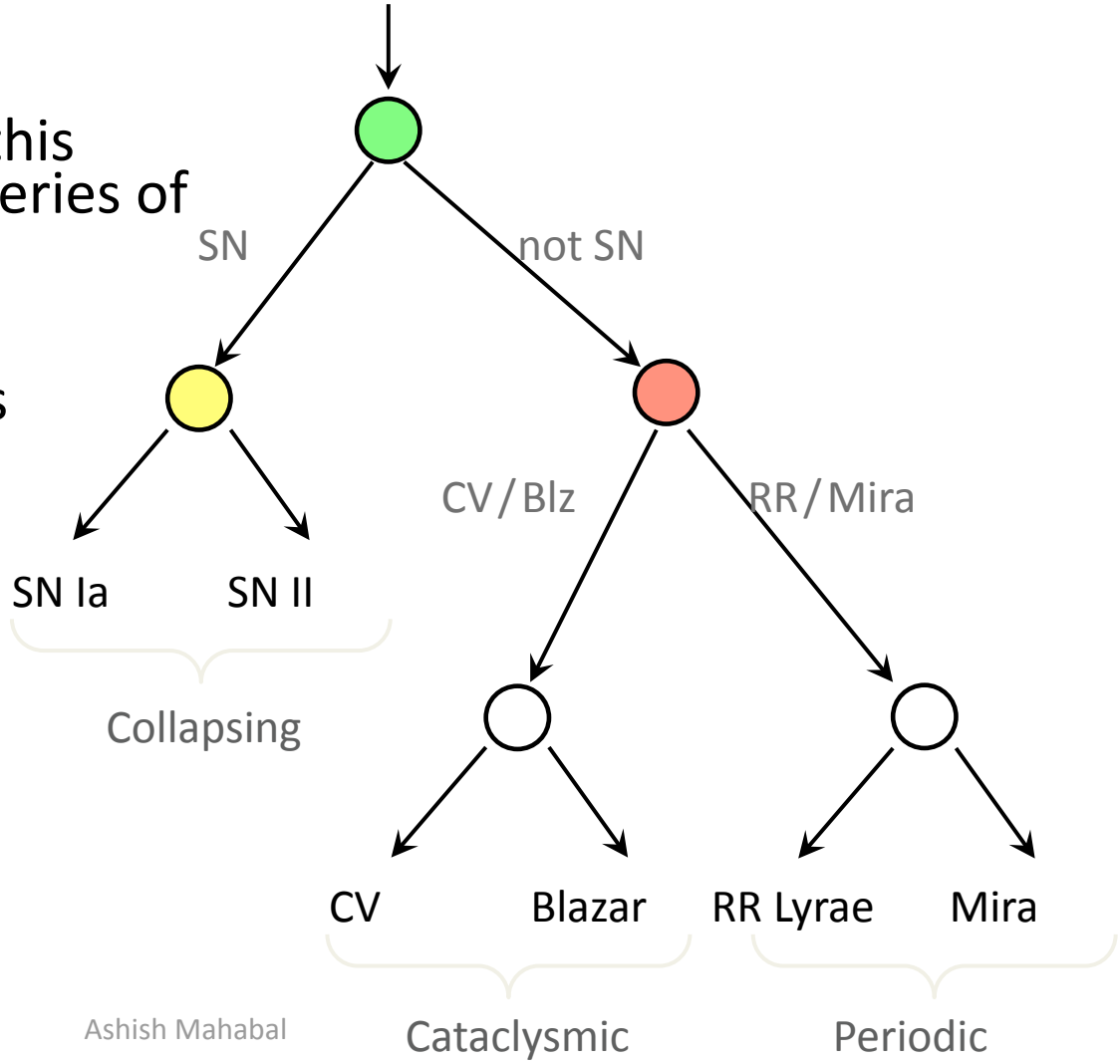
B Moghaddam, Y Chen

Classifier Architecture

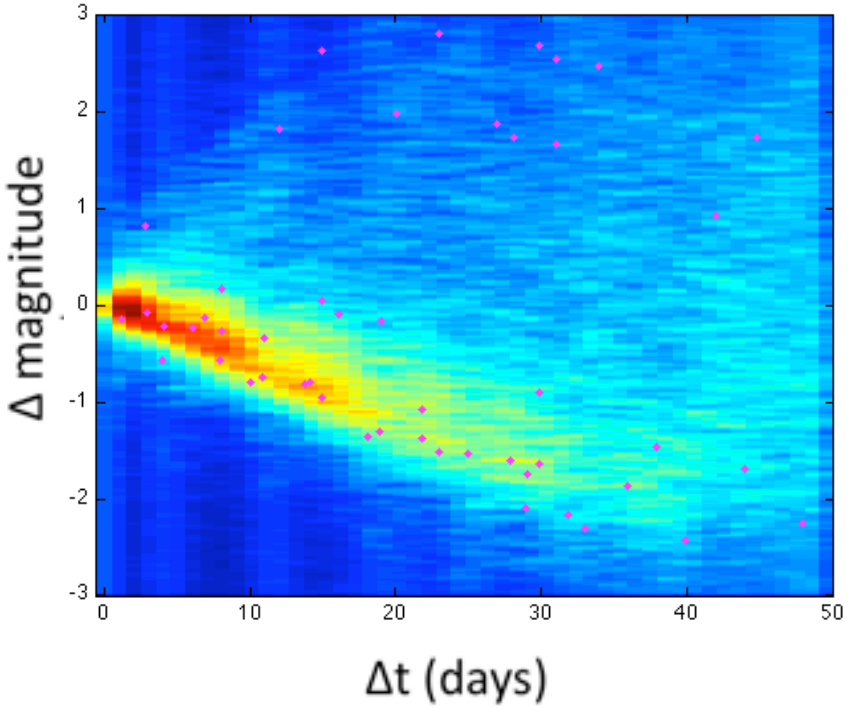
Decision Tree decomposes this multi-class classifier into a series of binary discrimination tasks.

This specific DT follows the stratification that seems natural to astronomers.

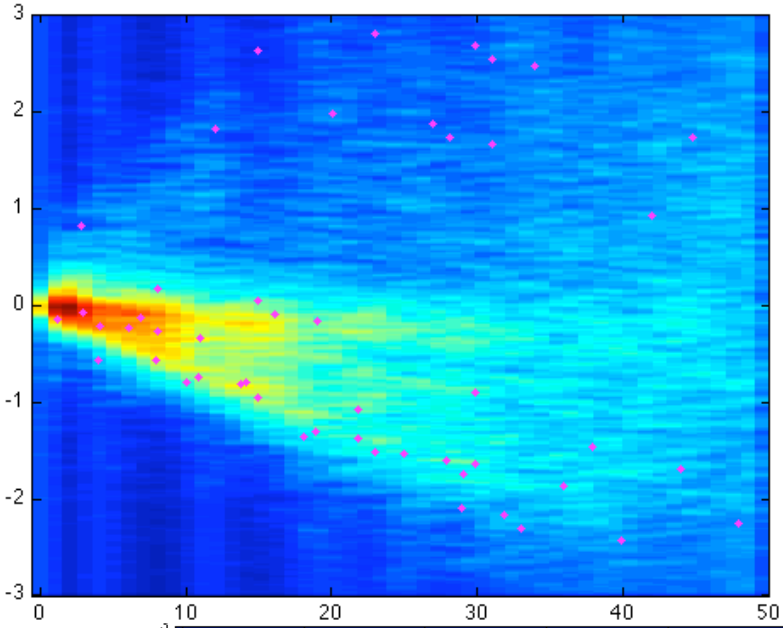
All nodes shown were Implemented via "gap histogram" binary classifiers.

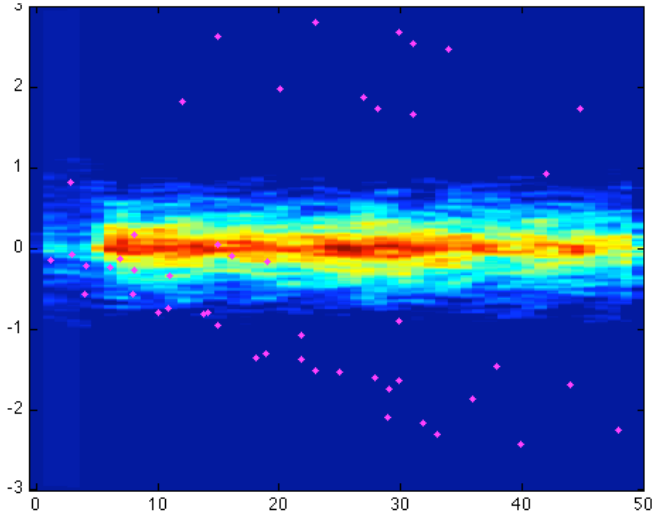


SN Ia



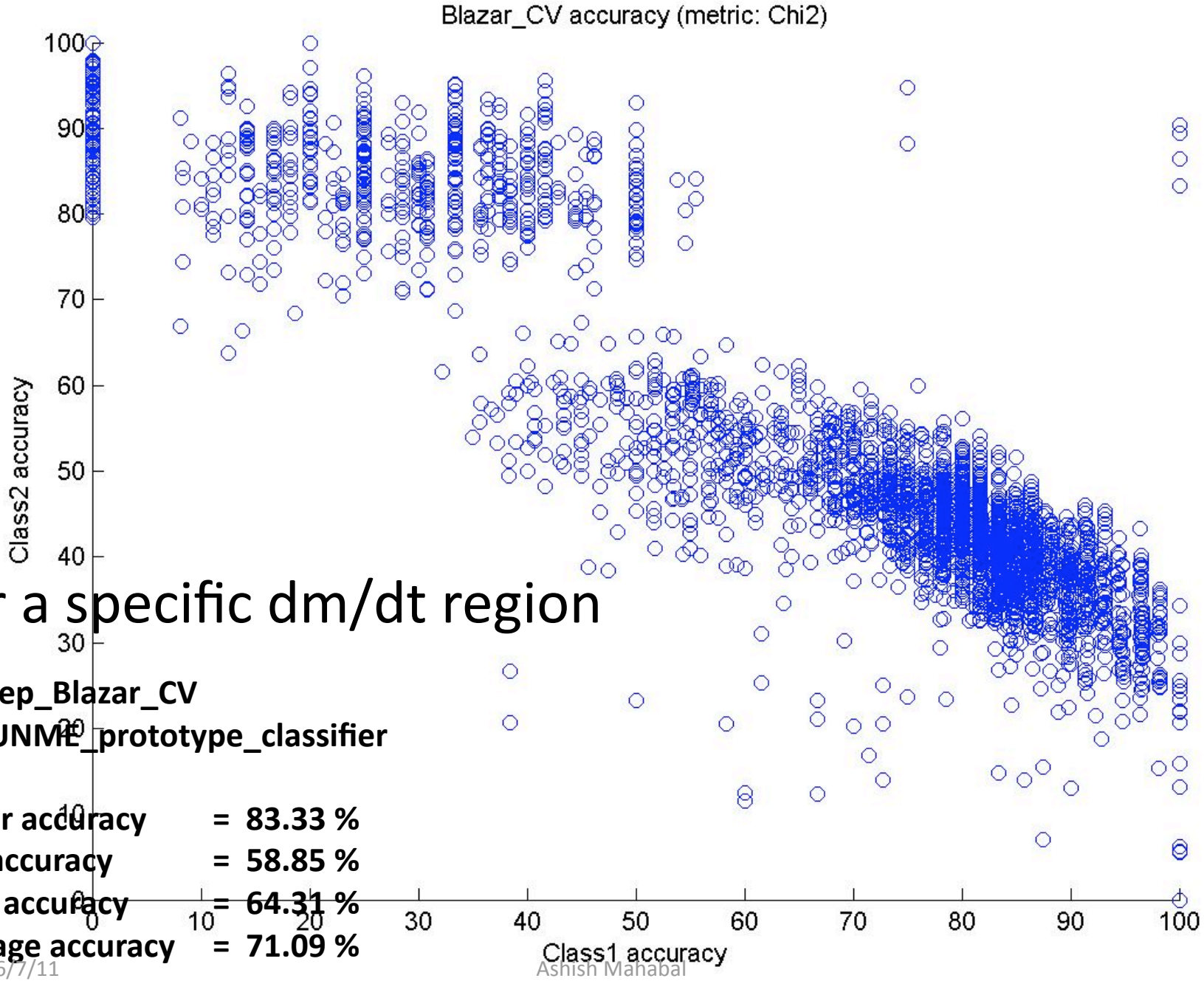
SN IIP





By taking subsections of dt/dm space determine which area is characteristic for which kind of variable

RR Lyrae



For a specific dm/dt region

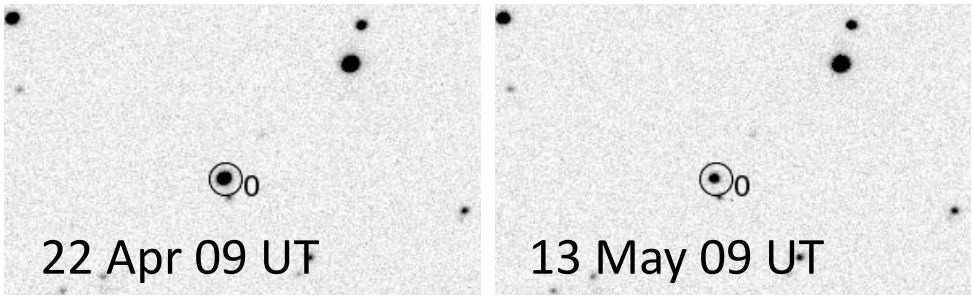
>> prep_Blazar_CV
>> RUNME_prototype_classifier

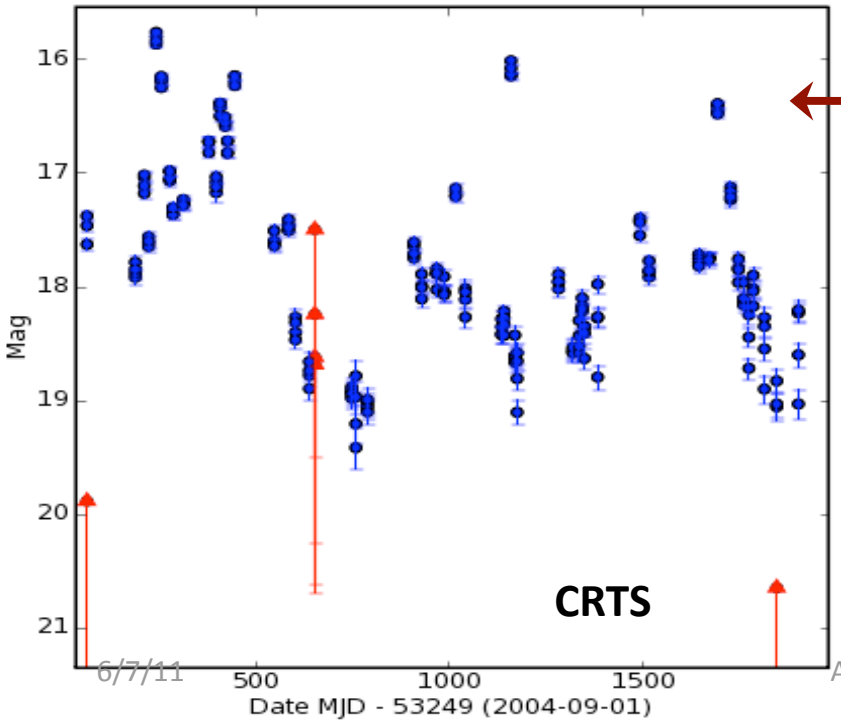
Blazar accuracy = 83.33 %
CV accuracy = 58.85 %
Total accuracy = 64.31 %
Average accuracy = 71.09 %

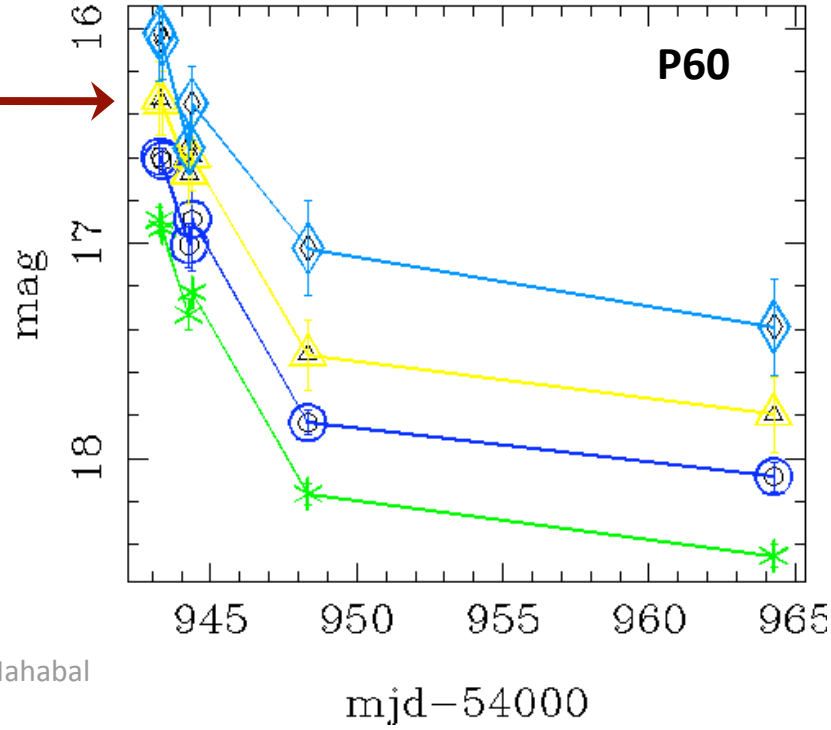
Follow-Up Observations:

- Photometry (P60, NMSU, DAO, HTN, India, Mexico, etc.)
- Spectroscopy (Gemini N+S, Keck, P200, SMARTS, IGO, MDM)

CSS090421:174806+340401 A blazar, also monitored at OVRO in radio



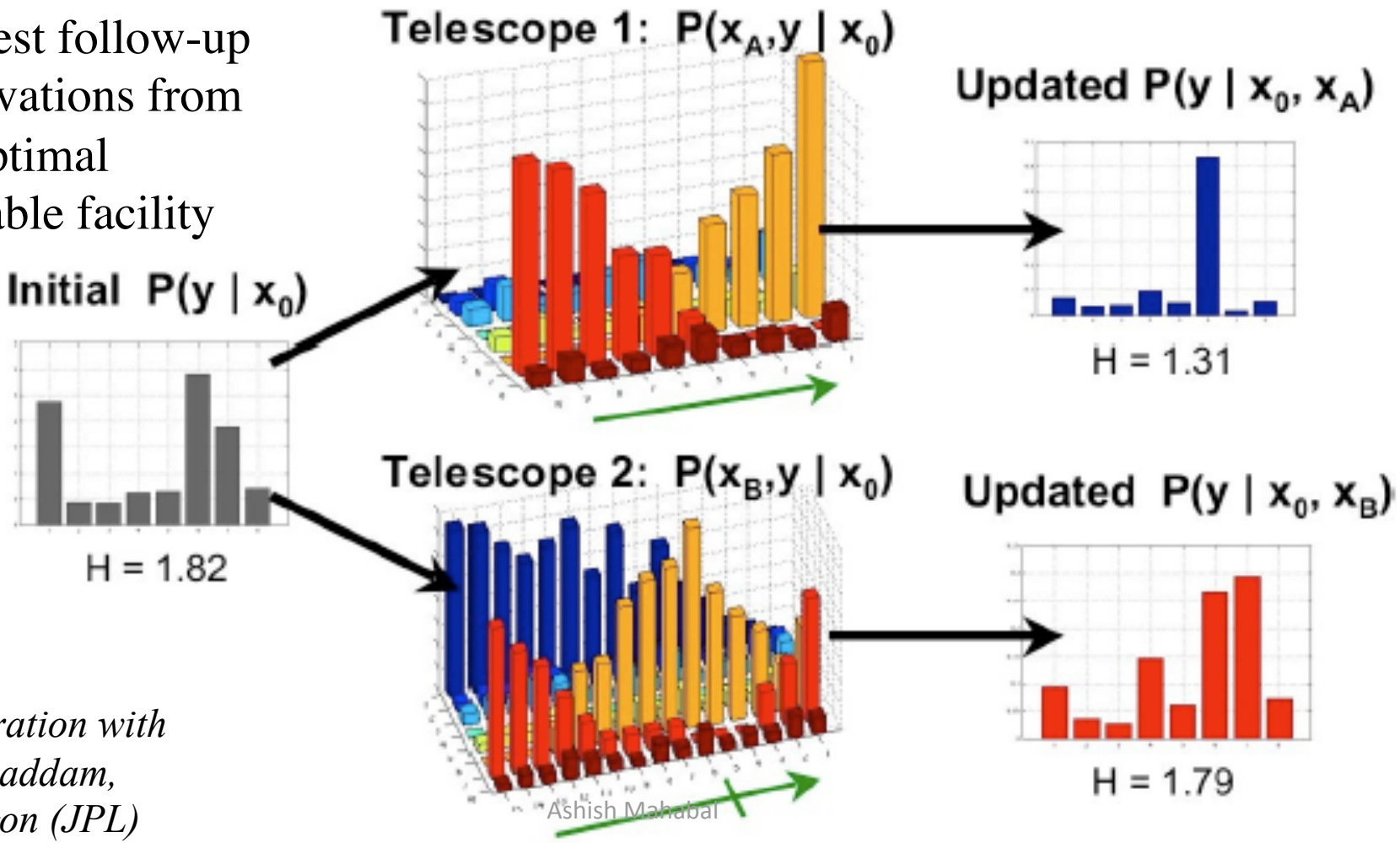




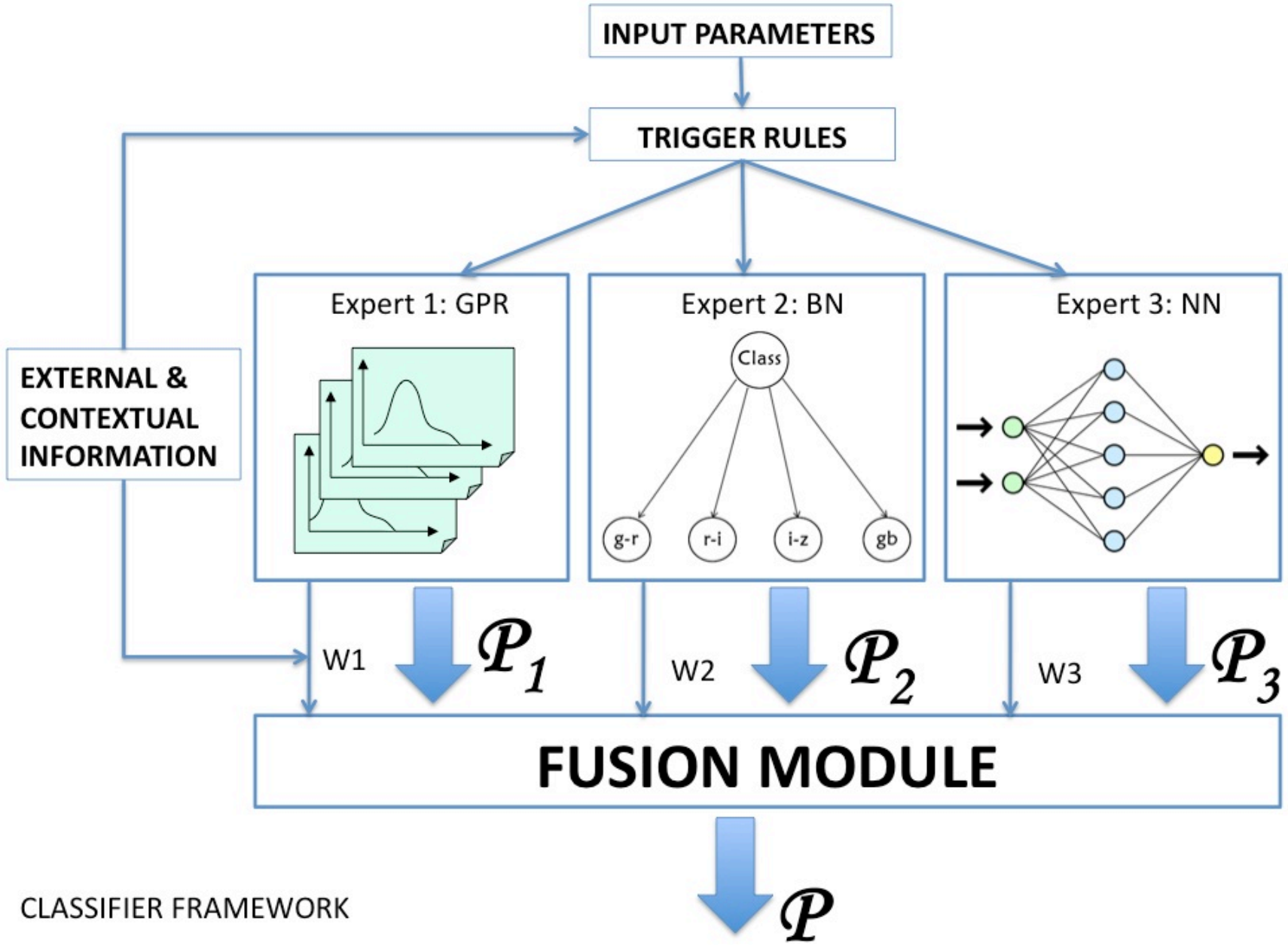
Automating the Optimal Follow-Up

What type of follow-up data has the greatest potential to discriminate among the competing models (event classes)?

Request follow-up observations from the optimal available facility



Collaboration with
B. Moghaddam,
M. Turmon (JPL)



CLASSIFIER FRAMEWORK

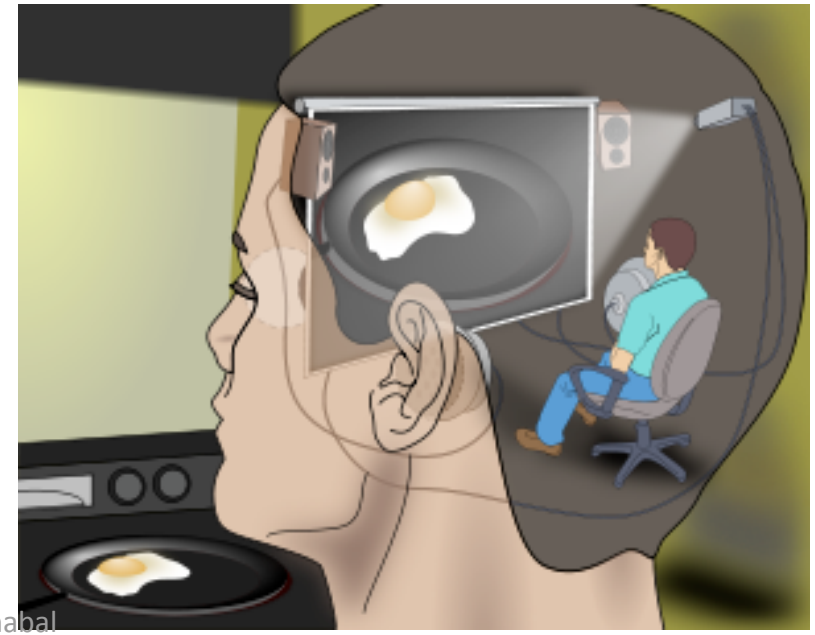
Bayesian Network/fusion modules are no Cartesian theatre

- Different parameters, methods are separate (though perhaps not independent) probes

(non-)Cartesian theatre

One observation can drive the direction given the large number of possible candidates

Not much scope for error



Event Publishing / Dissemination skyalert.org

PI: R. Williams

- Real time:
  - VOEvents, Twitter, iApp (thousands of events)
  - Also on SkyAlert.org, feeds to the WWT, GoogleSky
- Next day: annotated tables on the CRTS website

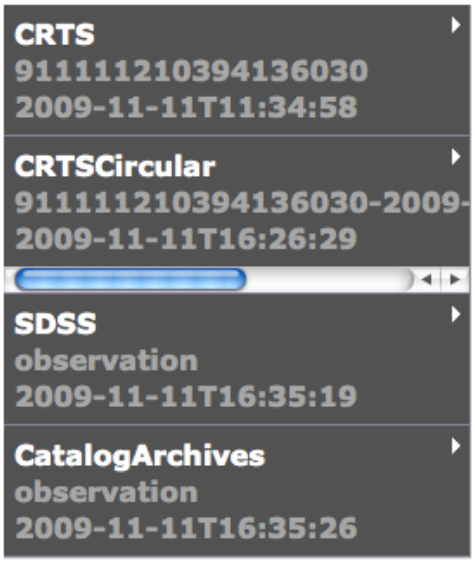
CSS ID	RA (J2000)	Dec (J2000)	Date	Mag	CSS images	SDSS	Others	Followed	Last	LC	Classification
CSS091121:221159+263906	332.99697	26.65153	20091121	18.33	911211261084134848	no	34848	no	2009-11-21	34848	SN/Blazar mag 21
CSS091121:013728+253450	24.36768	25.58061	20091121	17.78	911211260084103595	no	03595	no	2009-11-21	03595	SN/CV
CSS091121:032627+070744	51.61364	7.12902	20091121	16.68	911211070194124436	no	24436	no	2009-11-21	24436	CV mag 21
CSS091121:033232+020439	53.13295	2.07747	20091121	16.93	911211010194134434	no	34434	no	2009-11-21	34434	CV mag 20
CSS091121:085600-051945	133.99922	-5.32906	20091121	18.17	911210040484107252	no	07252	no	2009-11-21	07252	SN CFHT mag 22 gal
CSS091120:100525+511639	151.35223	51.27742	20091120	18.80	911201520354108835	yes	08835	no	2009-11-20	08835	SN SDSS mag 21,9 gal
CSS091120:082908+482639	127.28503	48.44423	20091120	15.69	911201490314109371	yes	09371	no	2009-11-20	09371	CV/SN SDSS mag 21,6 gal?
CSS091120:004417+411854	11.07004	41.31494	20091120	17.00	911201400044145995	yes	45995	no	2009-11-20	45995	Nova M31 2009-11d
CSS091120:001019+410455	2.58044	41.08191	20091120	16.69	911201400014137919	no	37919	no	2009-11-20	37919	CV mag 20,0

Real Time Event Publishing via *VOEvents* and *SkyAlert*

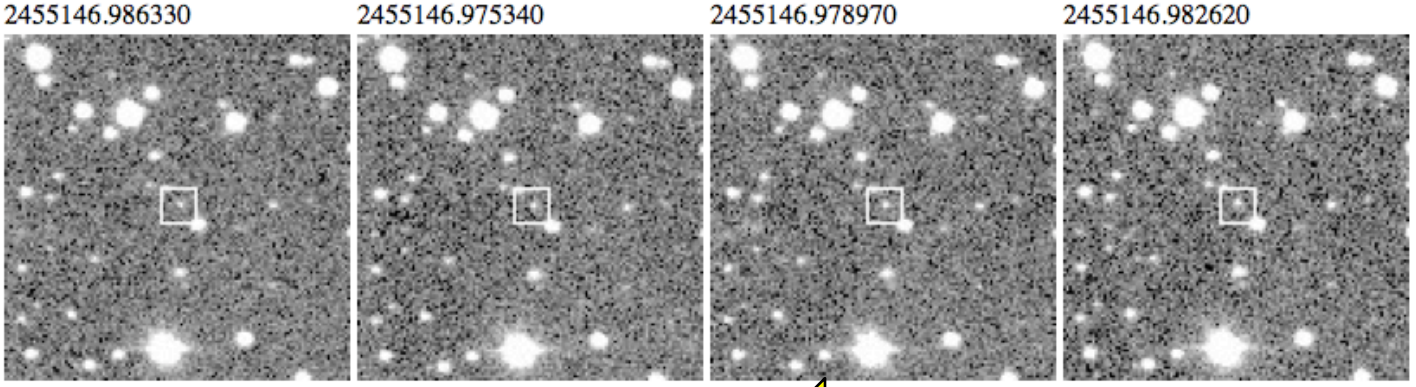
From the CRTS stream.
Catalina Real-time Transient Survey
Position is 115.98635,21.1753 ± 0.0012
This portfolio initiated 2009-11-11 08:35:18

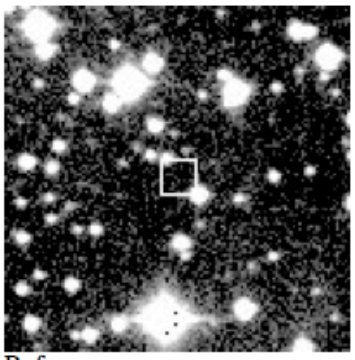
Basic event info

See context in WorldWideTelescope



CRTS (Catalina) Event identifier is 911111210394136030 or CSS091111:074357+211031





Reference

Finding Chart	Click here
Past CRTS images	Click here
Other images	Click here
Lightcurve	Click here
SDSS cutout	Click here
Position	(115.98635,21.1753)
Time	2009-11-11T11:34:58 (MJD 55146.4826157)
Magnitude	18.559
Magnitude	18.673

Linked VO/archival data for classif. and follow-up

Dynamically growing portfolio

Subscribe to VOEvents via email, RSS, Atom feed, etc.

http://www.skyalert.org/

- http://lib.skyalert.org/skyalert/Guide2Skyalert.pdf
- Subscribe to feeds – set-up alerts based on your own criteria
- Suppose you subscribe to a "CRTS + P60" alert
  - get the parameters that you want
  - be able to 'resolve' the event into its portfolio
  - wget –O myevent http://www.skyalert.org/params/ivo:++nvo.caltech+voeventnet+catot;90220018027414598 7/

Reading json output from skyalert

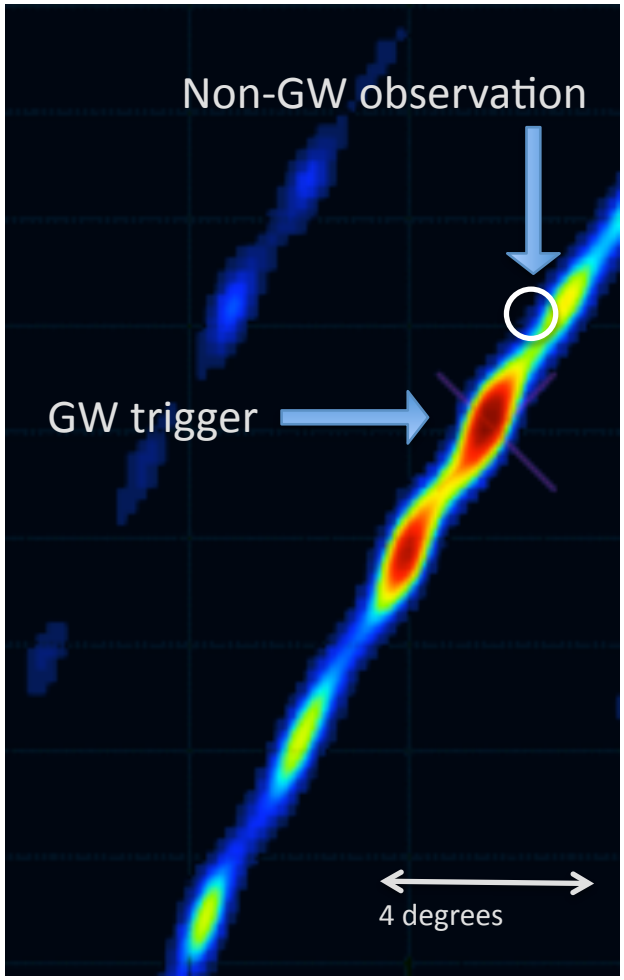
x <- fromJSON(file = ”myevent")
X

That gives:...
$portfolio$`ivo://nvo.caltech/voeventnet/csFollowup#observation`$nearest_obj_usnob
$portfolio$`ivo://nvo.caltech/voeventnet/csFollowup#observation`$nearest_obj_usnob$distance
[1] "0.5”

x$portfolio$`ivo://nvo.caltech/voeventnet/sdssFollowup#observation`[[1]]$ISOtimegives
[1] "2009-11-17T17:24:52"

Coincidence Machine

for multi-messenger astronomy



Courtesy A. Searle

Given a reported coincidence:
-- Gravitational wave trigger with time and probability density
-- Burst detection with time and position

Questions:
-- What chance these are physically connected?
-- Should we immediately slew telescopes?

DAME

DAME is a joint effort between University Federico II, INAF-OACN, and Caltech aimed at implementing (as web application) a scientific gateway for data analysis, exploration, mining and visualization tools, on top of virtualized distributed computing environment.

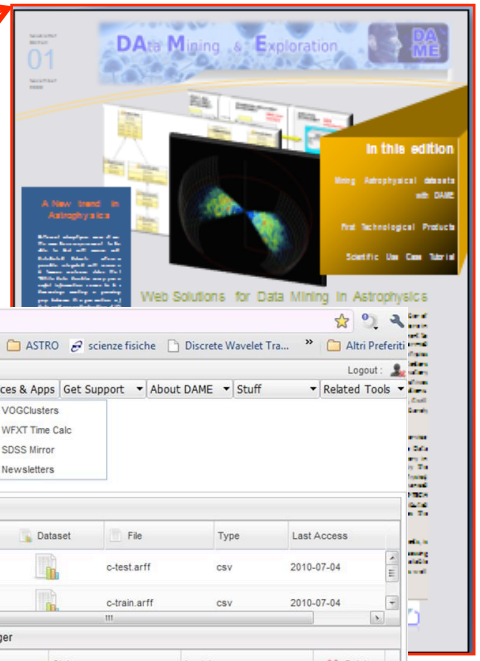
http://voneural.na.infn.it/

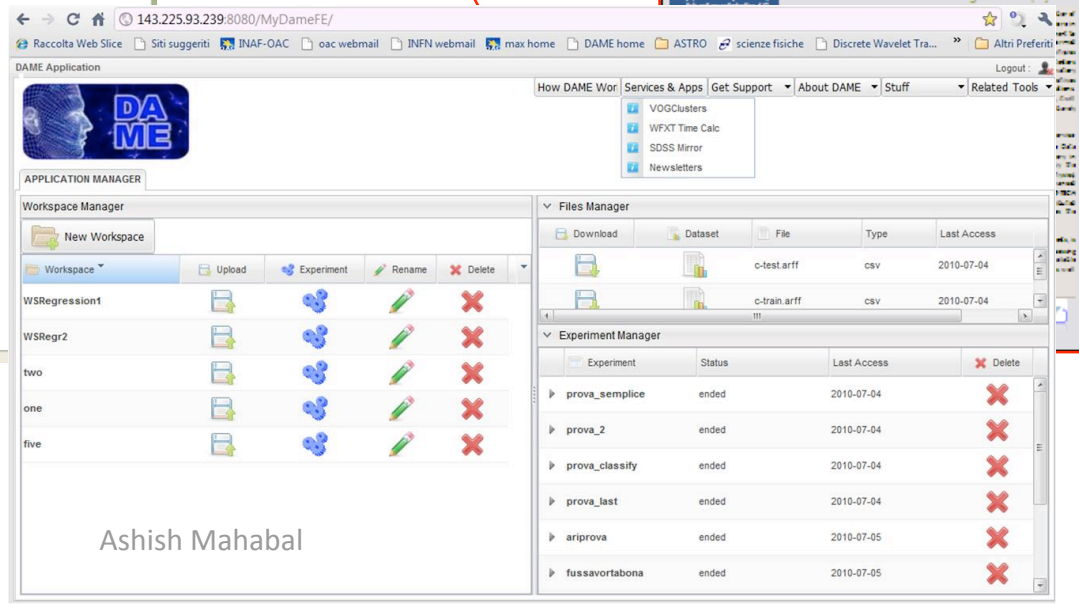
Technical and management info

Documents

Science cases

Newsletter





http://143.225.93.239:8080/MyDameFE/

Web application PROTOTYPE (ALPHA release)

brescia@na.astro.it

VOStat

Statistical Analysis for the Virtual Observatory

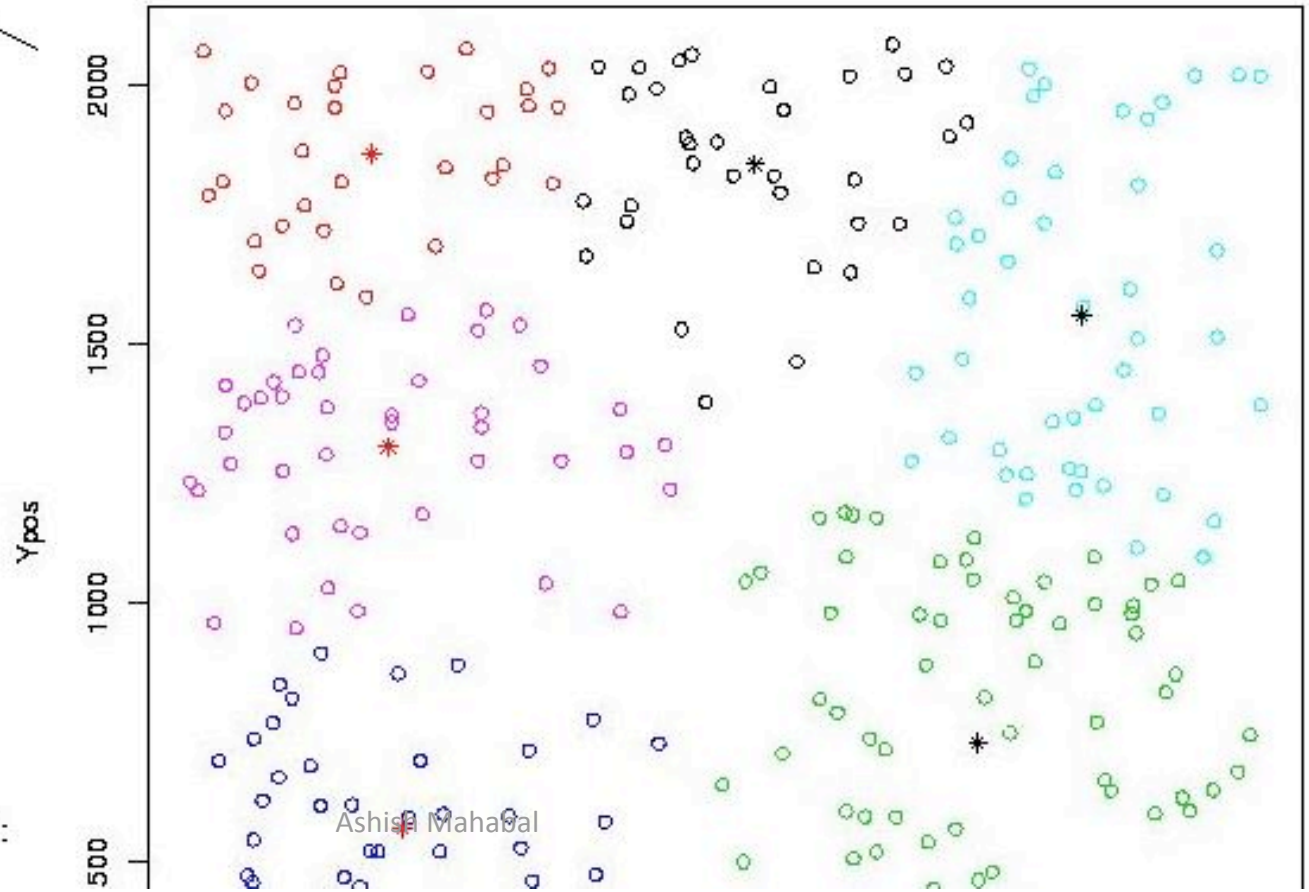
R Code | Plots | Save | Test List | Help | View File | View Data | VO Plot

RESULT

K-MEANS PARTITIONING

Click here to view Plot

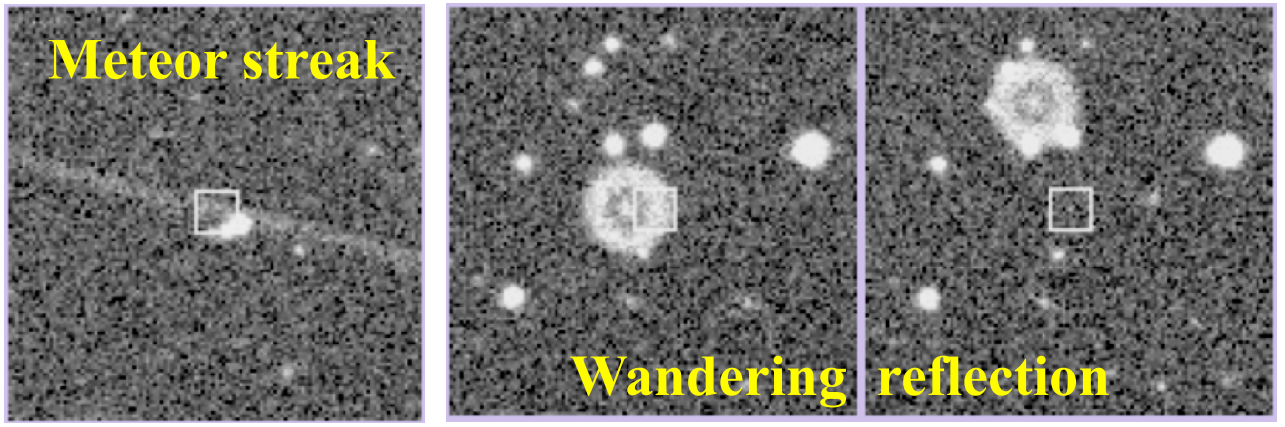
```
 sizes :

 33 30 73 42 46 42

 centers :

       Xpos      Ypos
1 1186.9091 1846.9091
2  514.2000 1866.8000
3 1578.8219  730.3562
4  568.3333  565.4286
5 1760.1304 1554.7174
6  542.5000 1301.9048

 withinss :

  2292505  1764038 10275335  35:
```



Harvesting the Human Pattern Recognition

Recognizing the artifacts (false transients)

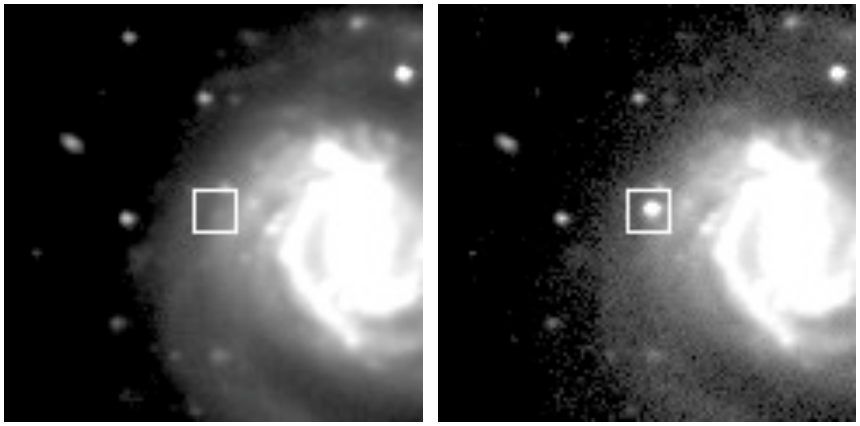
Contextual information is essential



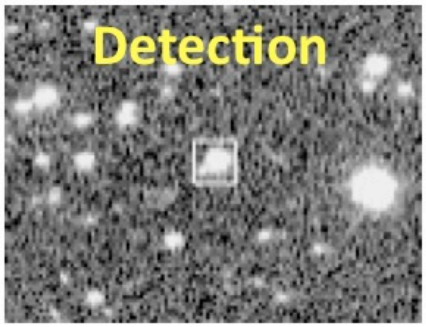
A more sophisticated case uses a **prior (expert) knowledge:**

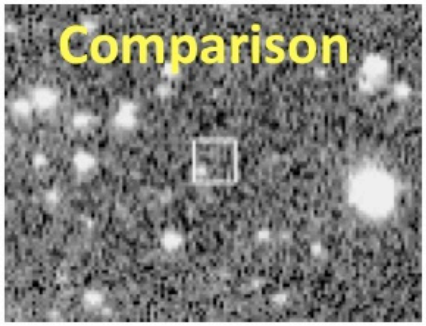
Star-like transient apparently associated with a non-coincident galaxy a likely Supernova

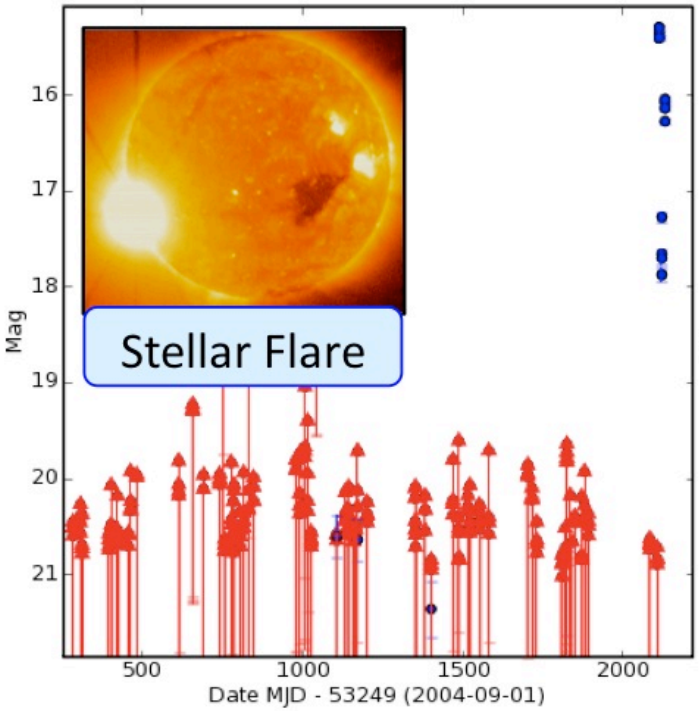
Spiral host galaxy
a possible Type II

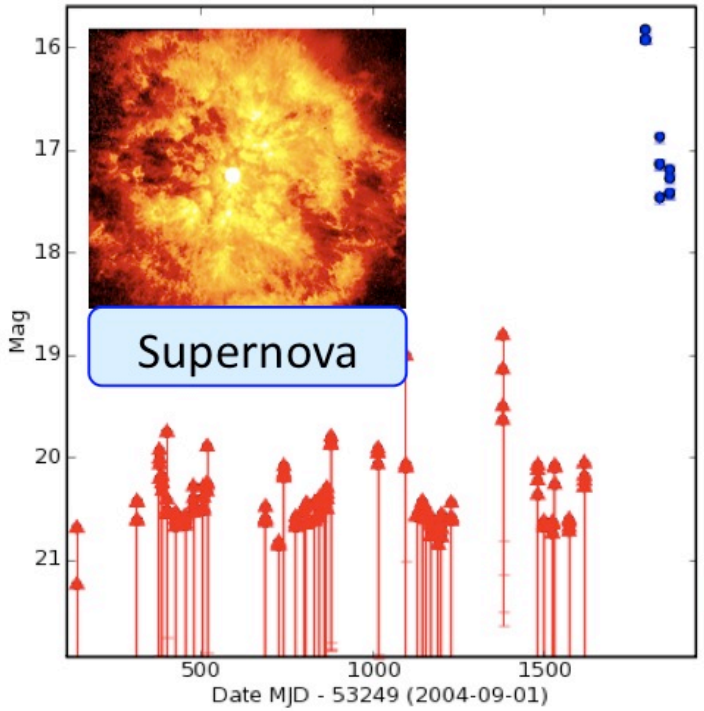


How to capture this and teach a machine to do the same thing?



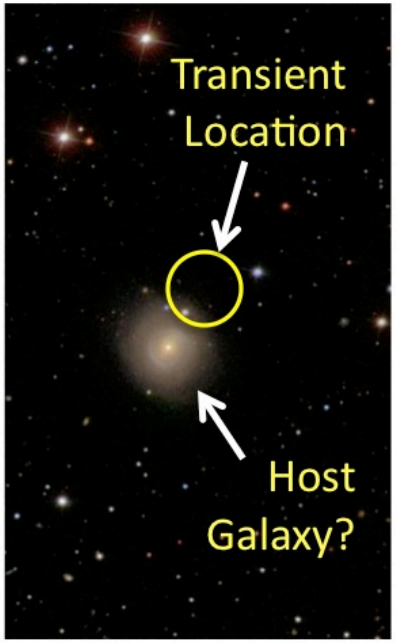




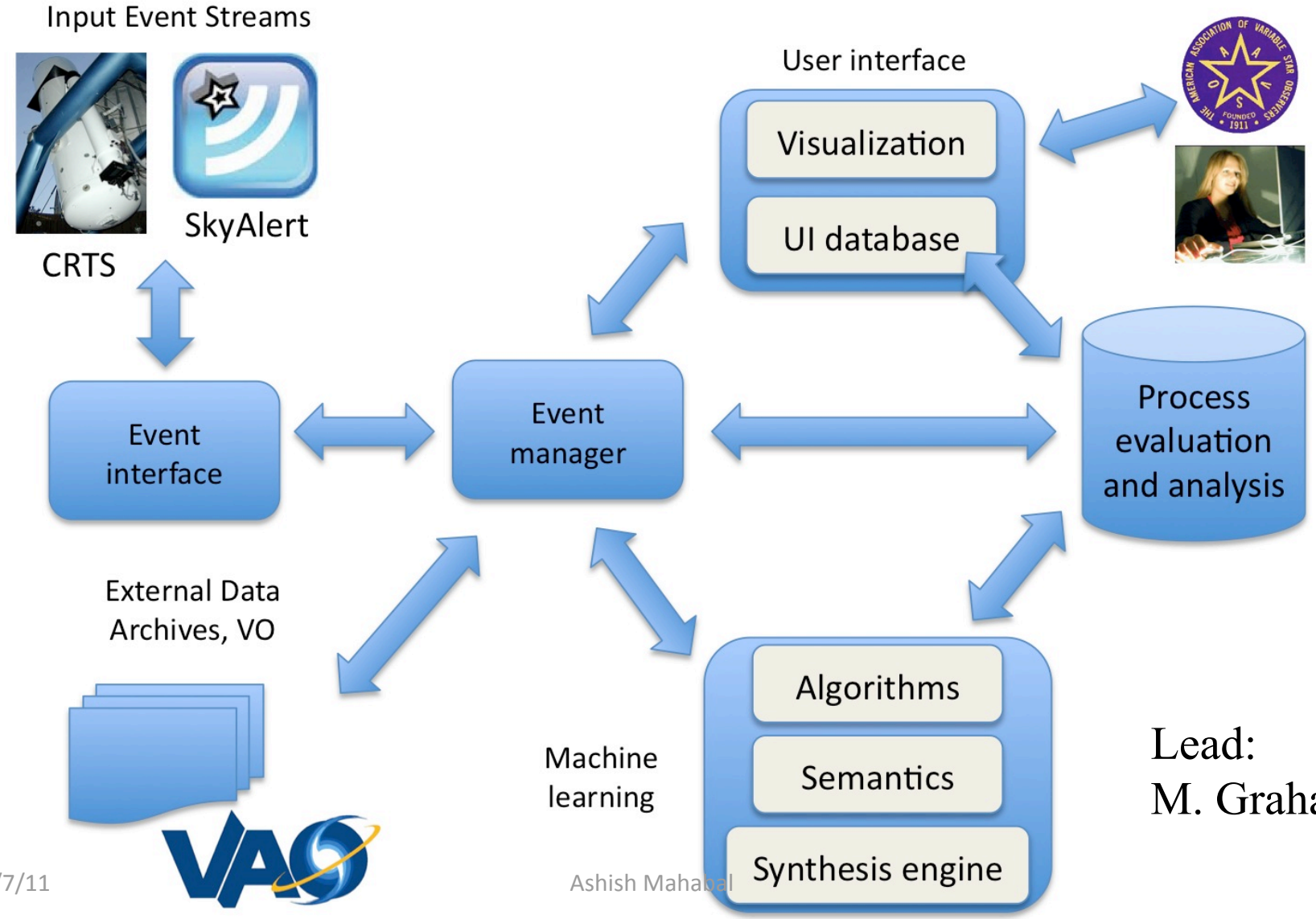


Use Case Scenario:

Light curves are ambiguous, but the presence of a possible host galaxy suggest that it is a Supernova



AstroCollation: Towards Harvesting Human Pattern Recognition and Domain Expertise



Lead:
M. Graham

Transient classification mantra

- Obtain a couple of epochs in one or more filters
- Assigns probabilities for different classes
- Choose observations (filters, wavelengths) for best discrimination
- Feed the new observations back in
- Revise probabilities, choose observations, …
- Based on confirmed class revise priors

Bayesian network, Gap processing, (DAME, VOStat, VO), Skyalert

Summary

Future: Minority Report like interfaces in open sims.

Video pictures ala sixth sense: Portfolios of transients (or any object for that matter) – automagically updating lightcurves, SEDs etc.

The Glass Bead Game of astronomy – connecting everything possible for classifying transients in realtime