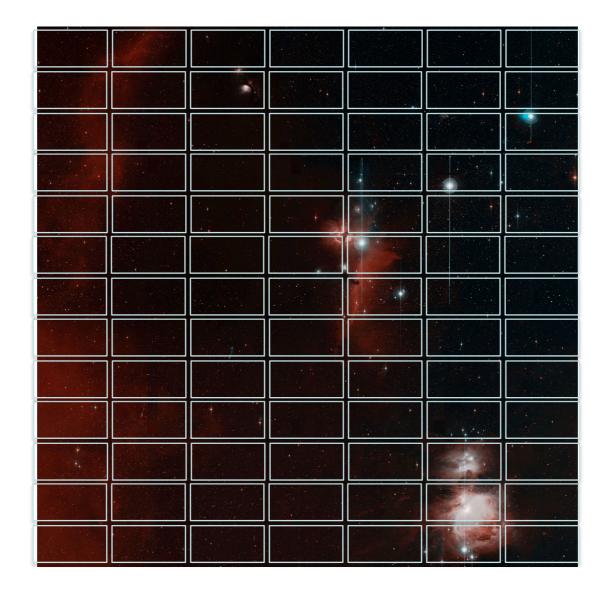
ZTF and Gaia synergy for classification of sources





Ashish Mahabal, aam at <u>astro.caltech.edu</u> ZTF ML lead, astronomy Center for Data-Driven Discovery 9th Gaia Science Alerts Workshop Vipava, Slovenia, 2018-10-08



Outline

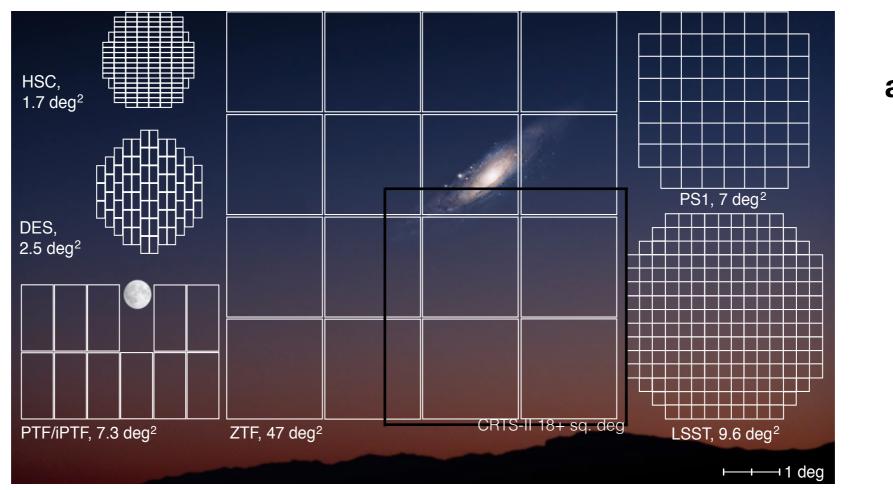
- Intro to Zwicky Transient Facility (ZTF)
- Data processing and science
- Connection to Gaia
- Possibilities

PEOPLE CLOSEST TO INSTRUMENTS AND DATA ARE BEST TO WORK WITH EVEN IF THE DATA ARE PUBLIC

Team Leads

PI	Shri Kulkarni
Co-PI	Thomas Prince
Project Scientist	Matthew Graham
Survey Scientist	Eric Bellm
Project Manager	Richard Dekany
Lead Camera Engineer	Roger Smith
P48 Operations	Tom Barlow
Data Archive Director	George Helou
Science Data System Lead	Frank Masci
Machine Learning Lead	Ashish Mahabal

ZTF has a large FOV and fast cadence



300K - 1M alerts/night

1000 observations over 3 years 1 billion sources First trillion row survey ZTF has bigger FOV (47 sq. deg.), new camera (600 M pixels), new CCDs, Nightly data: 1.4 TB more science use-cases

Different observing programs

- MSIP Public (g/r 40%) 3 nights (|b| < 7, every night)
- Partnership (g/r/i 40%)
- Caltech (20% mostly g/r)

MSIP - alerts public in real time Data to be made public Like with Gaia data we do not see/use non-public MSIP data for science

> March 17, 2018 – start of survey operations June 4, 2018 – start of public alert stream

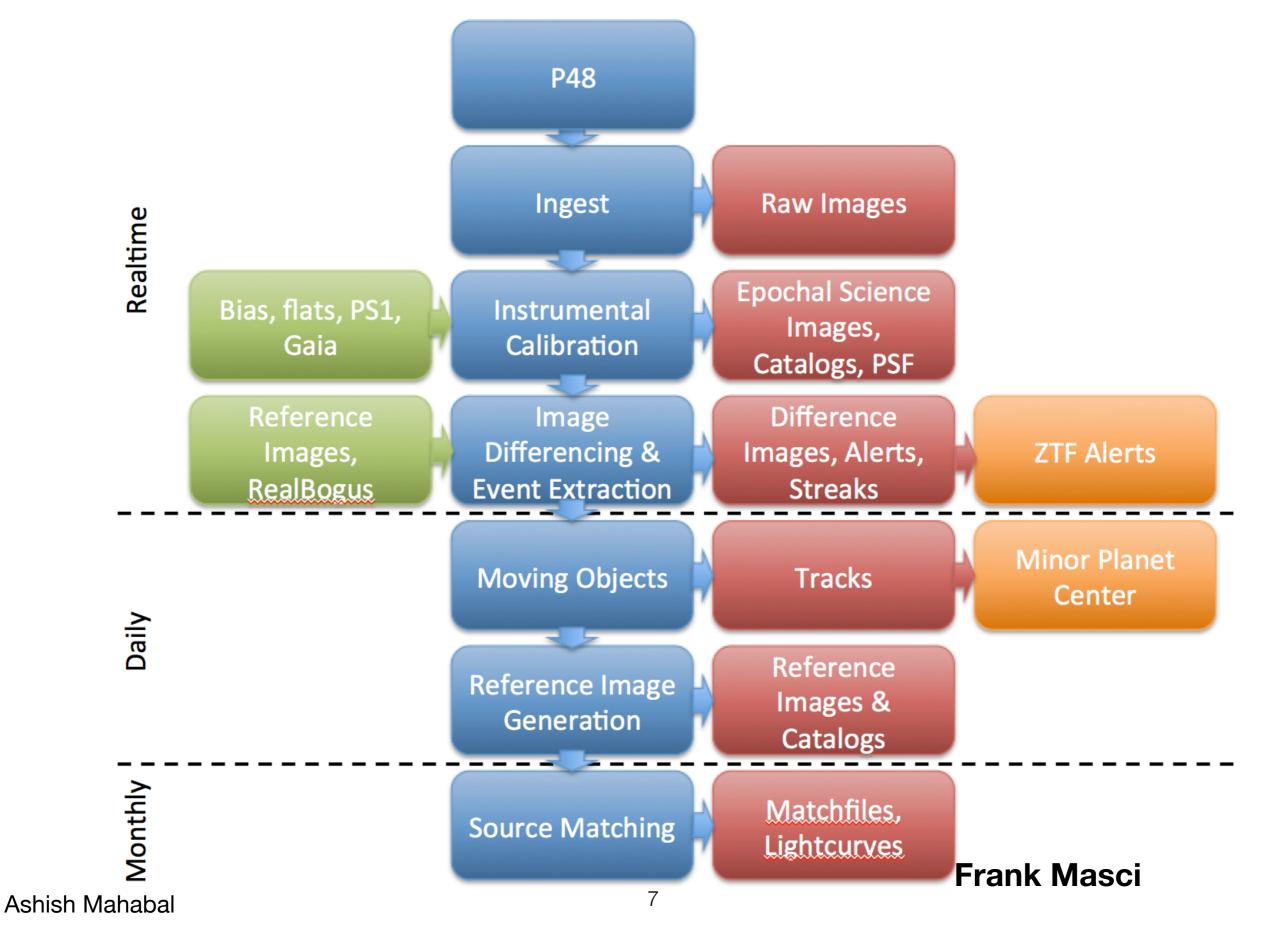
Caltech cadence

One of the programs (high cadence):

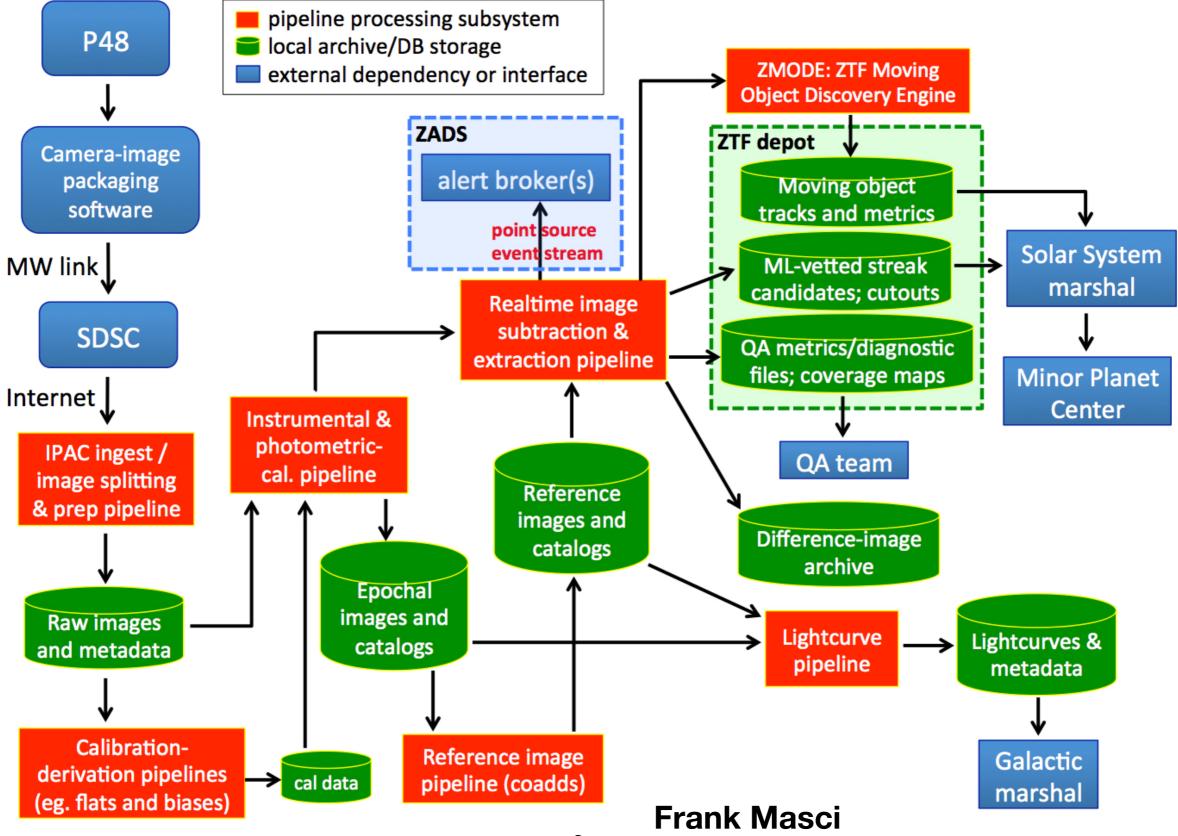
- -14 < GB < 14
- 90 minutes continuous
- repeated thrice (over few nights)

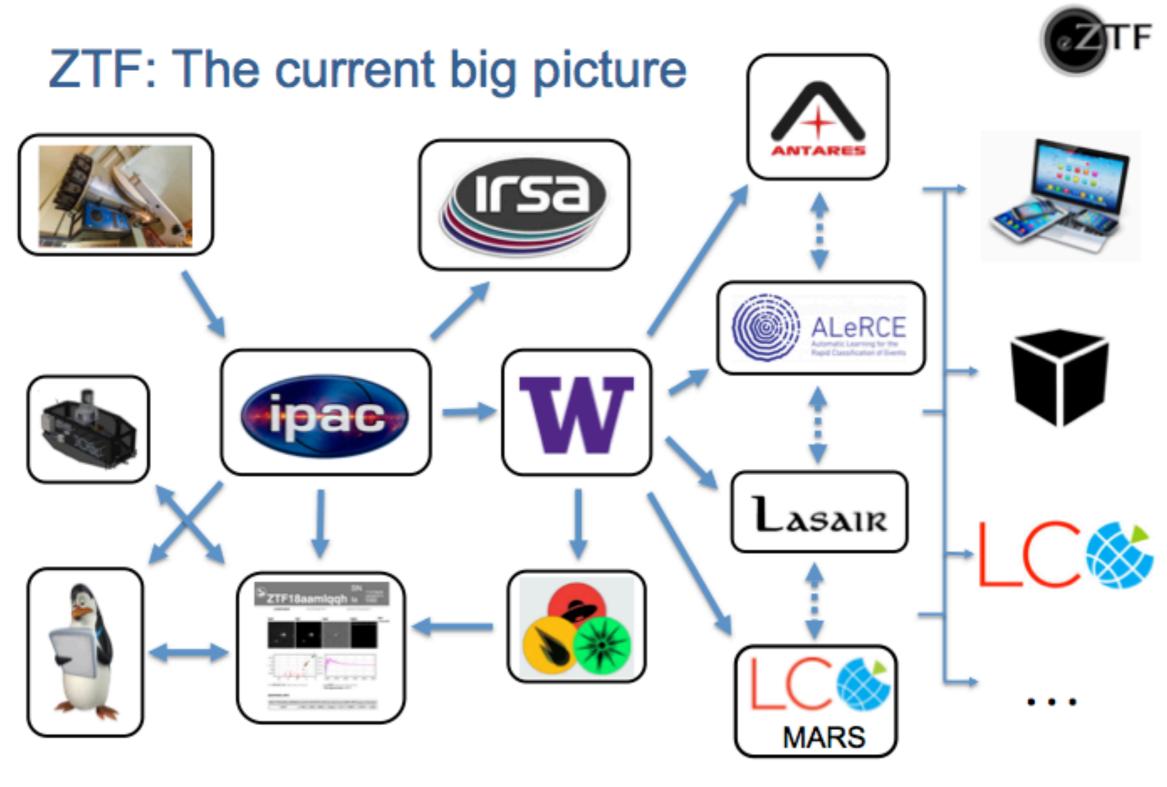
This is an area where we could do more together

All data uniformly processed - end points differ



For partnership data we have marshals







LASAIR



Lasair

Login

Coverage

Jupyter

Release Notes

Contact

Home	Query the ZTF database
Query	The form below is a builder for SELECT que
ZTF	 The attributes that can be queried are

Submit Query

magpsf < 15

FROM candidates WHERE

SELECT

be queried are listed in the ZTF schema. a = two filtors fid - 1 (a) and fid - 2 (r)The public our out up

• The public survey uses two filters: fid=1 (g) and fid=2 (r)		
• For each observing field of the survey and each of the g and r filters, ZTF will only issue candidate alerts when it has built up a reference image of that field with that filter, by stacking 15 good images.	Name	Search
 Once that is in place, each fresh image is subtracted from the reference, and any 5-sigma difference 	at a della d	01000011
generates a candidate alert.	ztī_public_2	20180804.tar.gz
• When a candidate is within 1.5 arcseconds of a previous candidate, it gets the same objectId. Thus a light	ztf_public_2	20180803.tar.gz
curve can be obtained from all the candidates that have a given objectId.		
 More details of the processing pipeline are available here. 	ztf_public_2	20180802.tar.gz

SELECT queries on the ZTF database of candidate alerts.

and rb >= 0.65 and nbad = 0 and fwhm <= 5 and elong

• Further cuts can be made to remove spurious candic' -- where it says "rb >= 0.65 and nbad = 0 and fwhm

(KXT	1 1 1
		Ma

ake Alerts Really Simple MARS provides access to all public alerts issued by ZTF since the start of the public alert stream on June 1, 2018. Subsets of the alerts, filtered by selectable constraints, may be identified and downloaded, either through this webpage or using the underlying API. Alerts are ingested as they are generated by the ZTF survey and are made available immediately, which is reflected by the "Latest Alert" value below. Users are advised to limit their request frequency to a reasonable time period, preferably allowing at least 5 minutes between requests. In addition to our own help page, users should refer to the ZTF website and the ZTF Alert Archive for documentation on ZTF and the generation of alerts.

The following table lists ZTF alerts in descending order by JD. Use the filters on the right to narrow down the results to interesting candidates. When the results look good, add ?format=json to the url. You can now access this url to retrieve the full data and use it in your scripts. You can access an alert's previous alerts by visiting /<id> in the json view or by clicking the id link in the table.

See the help page for descriptions of the table values and available filters.





Select	Prev	Next Results	: 5841854	Pages: 8	58419		-	atest A.)7:20:49		8-08-04	
Reset	id	objectId	time	filter	ra	dec	magpsf	magap	distnr	Δ maglatest	Δı
Sort By	6121378	ZTF18ablonfj	2018-08- 04	g	269.91461	66.96017	18.26	18.76	6.367		
time 🗘			04 07:20:49								
Sort Order	6121764	ZTF18abloniq	2018-08- 04	g	265.89783	66.70604	18.78	18.98	5.880		
Descending \$			04 07:20:49								
objectId	6121943	ZTF18ablonlr		g	257.63087	68.18566	18.12	18.56	11.711		
candid			04 07:20:49								
	6121964	ZTF18aakvzfc	2018-08- 04	g	256.65589	67.64165	18.00	18.26	0.588	0.12	-2
time (lower)			07:20:49								
mm/dd/y	6121989	ZTF18aaoechd	2018-08- 04	g	273.53155	70.73565	16.75	16.84	0.452	-0.06	-1
time (upper)		10									



Below you will find compressed tar archives of ZTF event alerts (observations detected in image differences). Each tar file contains alerts collected in the given night (UTC-based), with each alert stored in a separate file in the AVRO format. To get you started, we offer a repository with few basic utilities for reading AVRO-serialized data, as well as an example Jupyter notebook.

Why this service?

We are providing this archive as a stop-gap measure until public event brokers capable of receiving and redistributing the ZTF alert stream come online. We expect this by the end of 2018.

Known caveats

• The data provided on this site is generated automatically, has not yet been optimally calibrated or fully characterized, and uses evolving algorithms. It is best suited for searches of relatively bright transients. • The files provided contain a full, unfiltered, 5-sigma alert stream.

DIRAC

Size

5.4G

3.3G

8.9G

6.8G

3.3G

3.1G

4.1G

2.5G

1.9G

2.5G

RS

Depending on your science case, you may wish to improve the purity of your sample by filtering the data on the included attributes such as the signal-to-noise ratio or the real-bogus score.

Last modified

10 hours ago

1 day ago

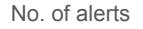
2 days ago

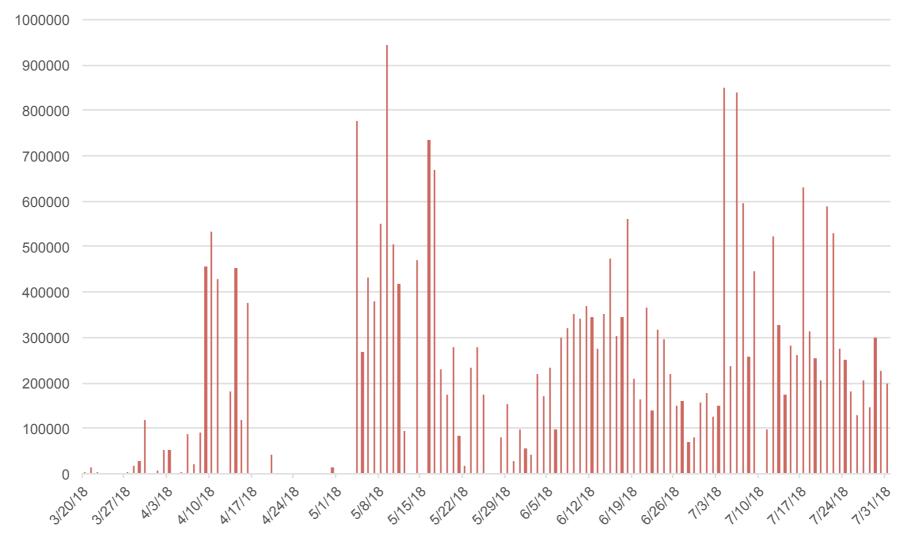
Help Contact

Ashish Mahabal	

MARS drilldown

lco_id	date	filter				
14494099	2458388.992037	r	Toggle JS9 M	enus		
lco_id	14912254		Science Dow	nload	Template Download	Difference Download
objectId	ZTF17aaarxib					
publisher	<pre>ZTF (www.ztf.caltech.edu)</pre>					
candid	643423150515010014				1	
	Candidate		1.0			-
jd	2458397.9231597					
wall_time	2018-10-06 10:09:20.998070		0000000			
fid	2		189216269376590	1013856555913	201 228 284 394 6151 052 92 367 97 150	-592588581-565535475356116359
filter	r					
pid	643423150515				Light Curve	
diffmaglim	20.4980869293213					
pdiffimfilename	ztf_20181006422998_000811_zr_c02	_o_q2_scimrefdiffimg	•	-		
programpi	Kulkarni			ł		
programid	1		19.5	1		
candid	643423150515010014		٥			MARS
isdiffpos	f		20 During			WIARO
tblid	14		nagr			
nid	643		_			т
rcid	5		20.5			
field	811					
xpos	2515.35522460938			Sep 28 Se	ep 30 Oct 2 Oct 4	Oct 6



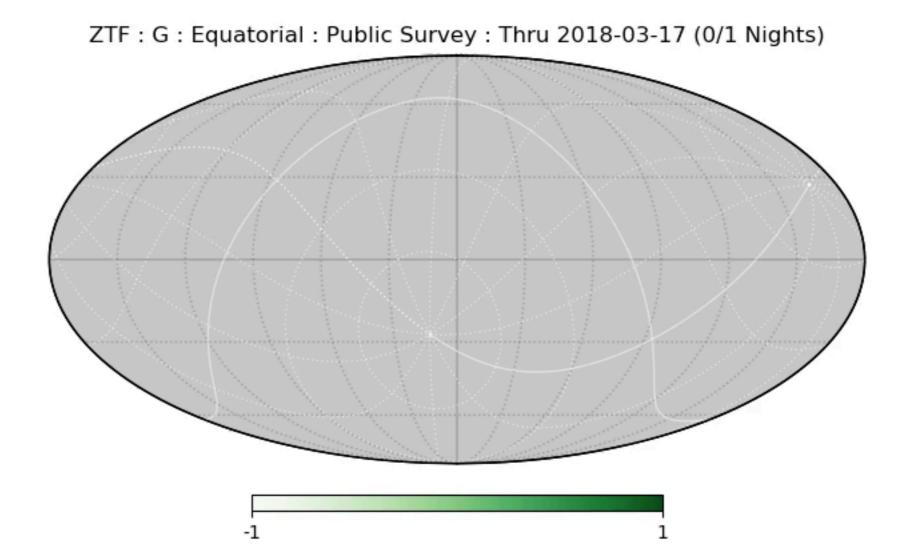


Astrometric reference sources from Gaia DR1 already in use for calibration. DR2 soon.

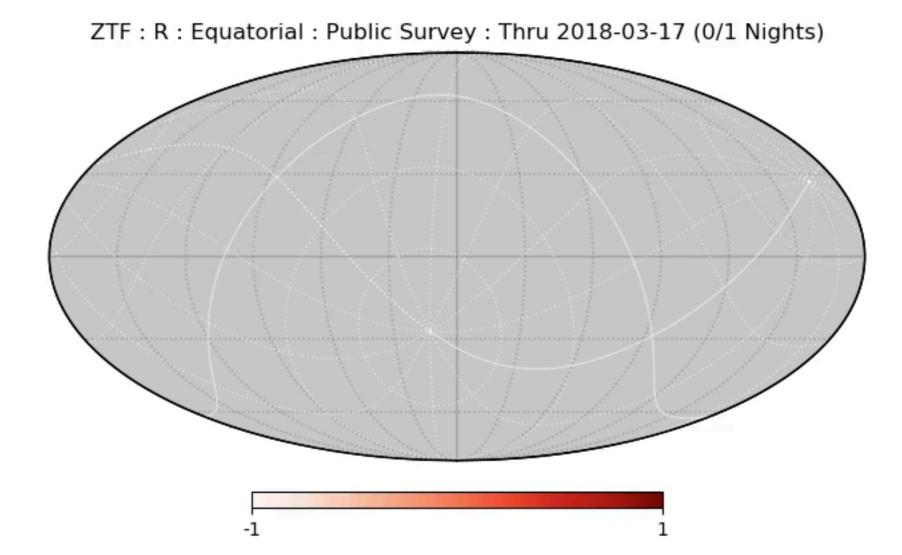
Gaia allows quantification of astrometric biases due to Differential Chromatic Refraction (DCR). DCR dominates above airmass of 1.8

Nearest Gaia DR1 source is reported in alert packets

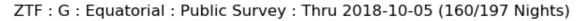
MSIP g band coverage

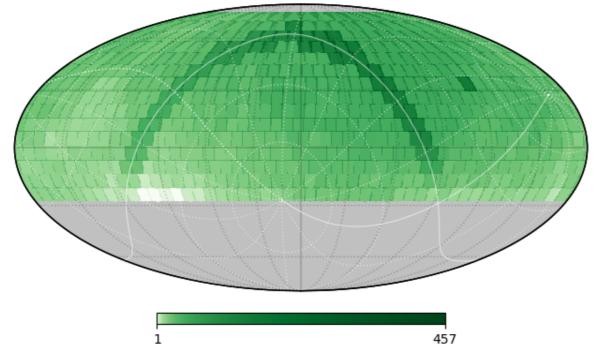


MSIP r band coverage

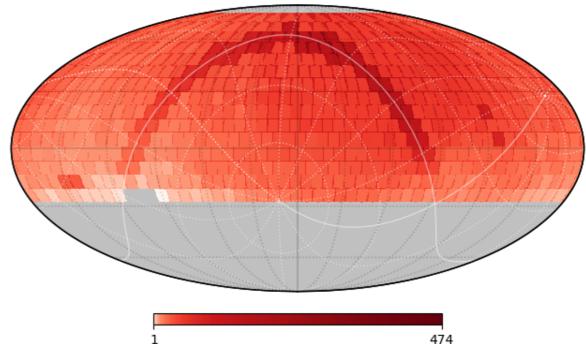


MSIP g and r band coverage

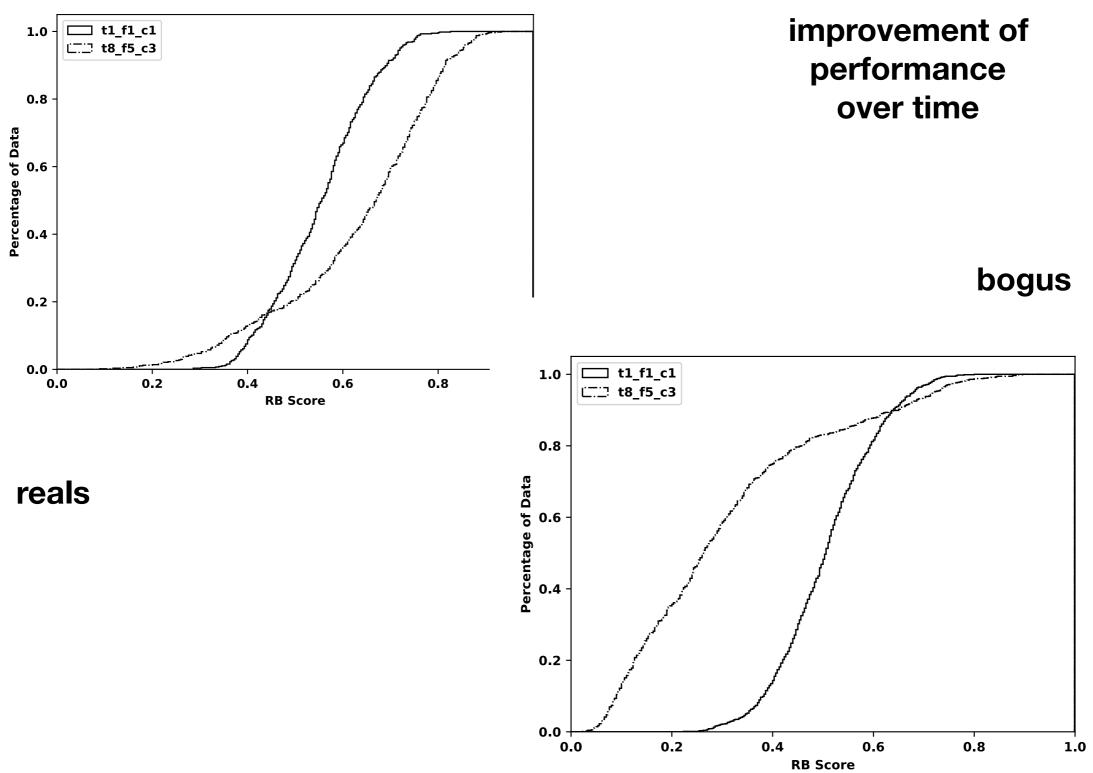




ZTF : R : Equatorial : Public Survey : Thru 2018-10-05 (166/197 Nights)



Real-Bogus separation



ZTF Science working groups

- Cosmology with Supernovae and Gravitational Lensing
 - Uli Feindt and Jakob Nordin
- AGNs and TDEs
 - Suvi Gezari and Matthew Graham
- EM/GW and Neutrino Counterparts
 - Mansi Kasliwal and Anna Franckowiak
- Physics of Supernovae and Relativistic Explosions
 - Ragnhild Lunna, Christoffer Fremling, and Steve Schulze
- Galactic and M31 Science
 - Thomas Kupfer and Po-Chieh Yu
- Solar System Bodies
 - Quan-Zhi Ye, Dennis Bodewits, and Rex Chang

Supernovae

> 400 la > 100 ll ~20 SLSNe >10 lb >10 lc

Lensed supernovae, catching all types on the rise, ...

First ZTF TDE

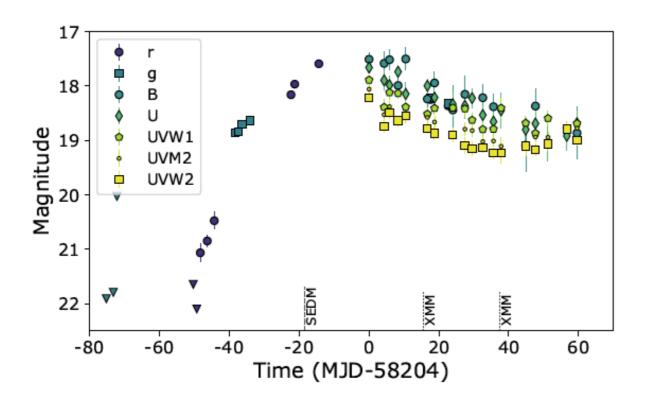


Figure 1. ZTF and *Swift*/UVOT light curve. The dashed line indicates the time of the first SEDM spectrum, while the dotted lines label the time of XMM X-ray observations. Triangles denote 5σ upper limits to the flux.

Van Velzen et al. 1809.02608

AT2018zr ZTF18aabtxvd ZTF-NedStark

z=0.071

x-ray detection but no radio

Gaia transients in galactic nuclei

1808.03984

Z. Kostrzewa-Rutkowska^{1,2*}, P.G. Jonker^{1,2}, S.T. Hodgkin³, Ł. Wyrzykowski⁴, M. Fraser⁵, D.L. Harrison^{3,6}, G. Rixon³, A. Yoldas³, F. van Leeuwen³, A. Delgado³, M. van Leeuwen³, S. E. Koposov^{7,3} ¹SRON, Netherlands Institute for Space Research, Sorbonnelaan 2, 3584 CA Utrecht, the Netherlands

² Department of Astrophysics /IMAPP Radboud University P.O. Box 0010 6500 CL Niemann the Natherlands

von Neumann and skewness

gorithm. We calculate the von Neumann statistic – the ratio of the successive mean square difference to the variance:

$$\eta = \frac{\frac{1}{n-1}\sum_{i=1}^{n-1}(x_{i+1} - x_i)^2}{s^2};$$
(1)

(von Neumann 1941) as well as the skewness of the light curve:

$$\gamma = \frac{\frac{1}{n} \sum_{i=1}^{n-1} (x_i - \bar{x})^3}{s^3}$$
(2)

(where x are the flux measurements during transits, s- the variance of the light curve, n - the number of transits in the light curve). These two statistics were

AlertPipe OldSource NewSource

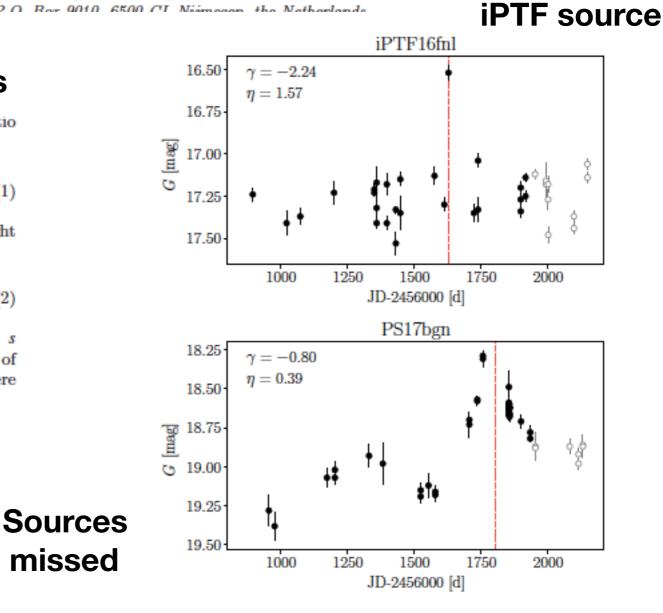


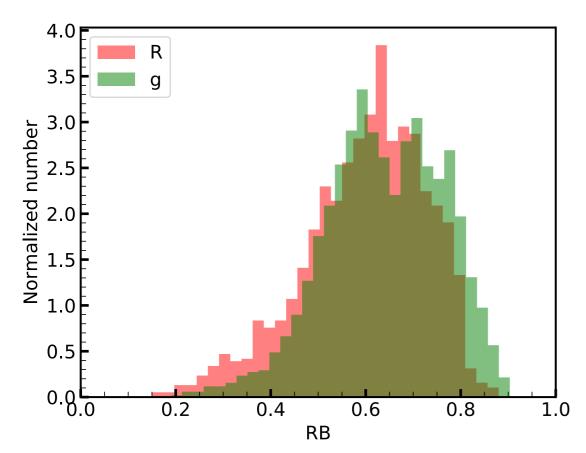
Figure 9. Examples of light curves of transient sources discovered by other surveys but not by Gaia even though the sources were detected. The dashed red line indicates the discovery time. Top:

Ashish Mahabal

20

missed

Comparison of ZTF sources with TNS



Mahabal et al 2018 Blagorodnova

Also validation through detection of known asteroids

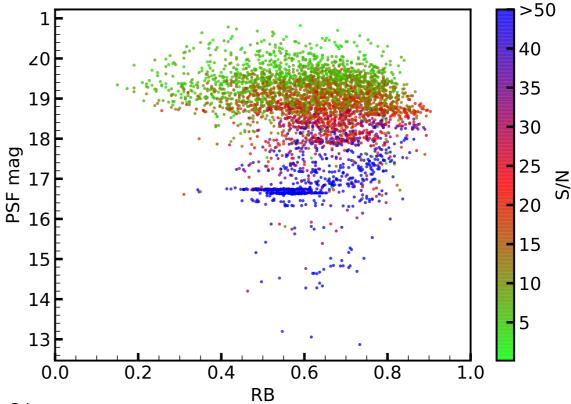


Figure 8. Top: RB score distribution in each band for ZTF alerts matched to TNS sources with discovery dates between UTC 2018 June 2nd to 2018 August 14th. Bottom: Scatter plot of the PSF magnitudes of the transients versus their RB scores. The color scale indicates the signal-to-noise ratio of the residual.

Wevers et al.

4

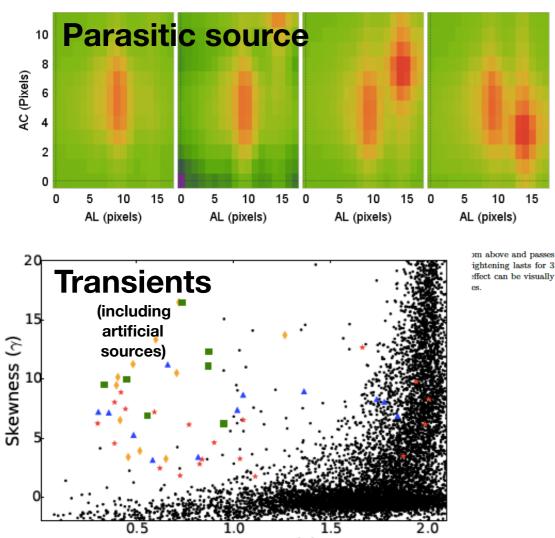
The fast transient sky with Gaia

1710.08924

Thomas Wevers^{1*}, Peter G. Jonker^{2,1}, Simon T. Hodgkin³, Zuzanna Kostrzewa-Rutkowska^{2,1}, Diana L. Harrison^{3,4}, Guy Rixon³, Gijs Nelemans^{1,5}, Maroussia Roelens⁶, Laurent Eyer⁶, Floor van Leeuwen³ and Abdullah Yoldas³

¹Department of Astrophysics/IMAPP, Radboud University, P.O. Box 9010, NL-6500GL Nijmegen, The Netherlands 2 about Maria 1 1 1 min 4 A ME OFFICIA DE LE TRE MELLE 1 0 1 .

15



von Neumann (η)

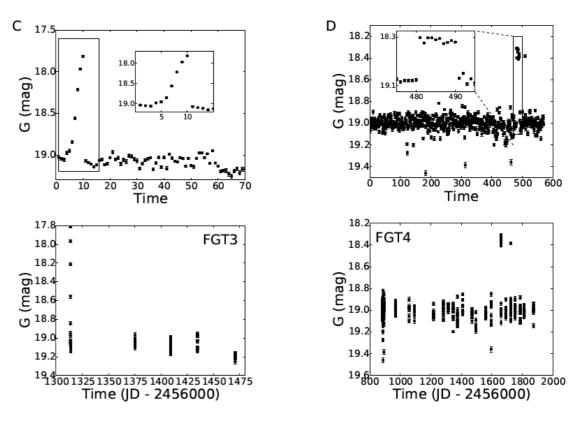
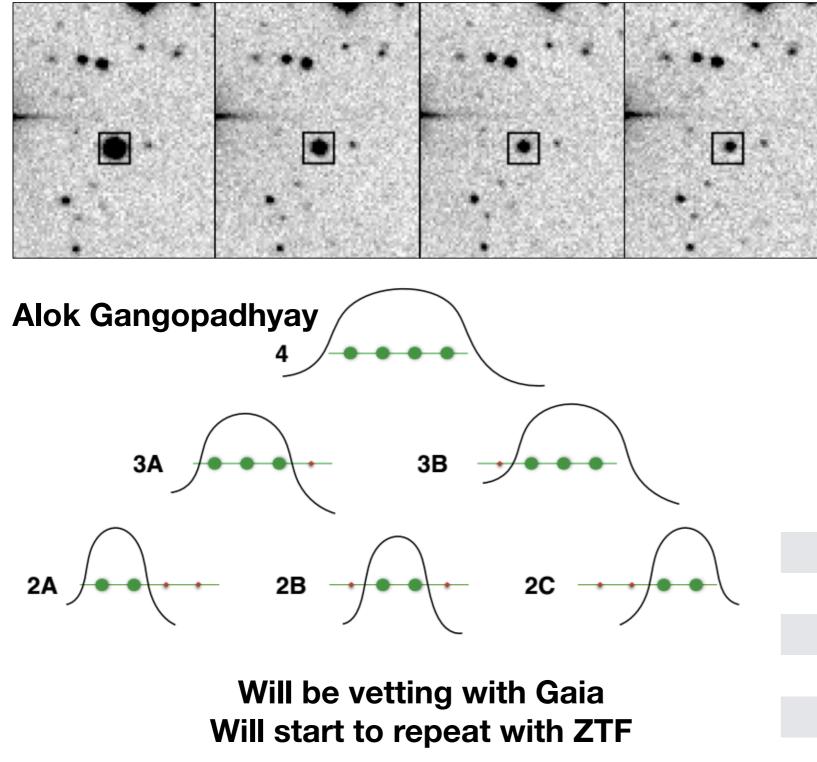


Figure 5. Lightcurves of FGT1 (panel A) through FGT4 (panel D). Each panel consists of 2 figures: the top figure shows the concatenated lightcurve for display purposes, with a running number per CCD crossing (i.e. time increases with running number). 10 measurements constitute 1 transit. The inset shows a zoom-in to the transient event. The bottom panel shows the actual time spacing of the observations, in Julian days. The data of the outlying transits is presented in tabular form in the Appendix.

A couple of the sources found

orphancat vetting through Gaia



Igor Anderoni

Photcat Orphancat CRTS nK transients 1.87B

	•	Ļ		;	•	•
ŧ						

	unfiltered	filtered
4-4	1.86 M	0.96 M
4	172 K	64 K
3A	198 K	78 K
3B	77 K	18 K
2A	735 K	75 K
2B	486 K	62 K
2C	643 K	79 K

Real-Bogus separation with large number of features

Image Level Features					
Limiting Magnitude	Expected 5-sigma magnitude limit of the sci and ref images after gain and background matching (and resampling for the reference image). Expected 5-sigma magnitude limit of diff image.				
Flux Ratio	Median flux ratio (Fluxsci / Fluxref) across matched sources.				
Sci Image	Electronic gain (small fluctuations are observed), saturation level (after gain-matching), modal background level and robust sigma/pixel after gain and background matching.				
Ref Image	Saturation level (after gain-matching) and resampling, modal background level and robust sigma/pixel after gain, background matching and resampling.				
Diff Image	Robust sigma/pixel, number of bad pixels before and after PSF-matching, percentage of pixels that are bad or unusable, effective FWHM in diff image, the average of squared diff image pixel values before and after PSF-matching, percentage of changed diff image pixels values before and after PSF-matching, status of image differencing (0=bad,1=good).				
Positive vs. Negative Diff	Median background level in positive and negative diff image, and number of candidates extracted from the positive and negative diff image before and after internal filtering.				
Signal-to-Noise Ratio (SNR) in Diff Image	Number of noise pixels in diff image, peak-pixel SNR in detection image optimized for point source detection, and ratio of the mean square pixel value in the subtraction image to its spatial variance following PSF matching.				
Seeing	Seeing of sci and ref images, and ratio of FWHM to average FWHM on the sci image.				
Candidate Features					
PSF Photometry	Magnitude and 1-sigma uncertainty from PSF fit; chi of candidate				
Aperture Photometry	Magnitude and 1-sigma uncertainty from "big" aperture photometry				
Candidate Measurements	Local sky background level, FWHM of local Gaussian profile, magnitude difference of PSF and aperture pho- tometry, magnitude difference of PSF photometry and limiting magnitude, and minimum distance to edge.				
Nearest Ref Source	Distance, magnitude, 1-sigma uncertainty, chi and sharpness of nearest reference image extraction.				
Nearest Solar System Object	Distance, magnitude of nearest solar system object				
Shape	Windowed RMS along major, minor axis of source profile, ratios of the major and minor axes to the FWHM, and elongation of the candidate.				
Negative/Bad Pixels	Number of negative pixels in a 7x7 box, number of bad pixels in a 7x7 box, ratio of sum of pixel values to sum of absolute values in source-stamp cutout.				

Larger number of features helps. Also exploring deep learning

Ashish Mahabal

Random Forest

Short timescale variables in the *Gaia* era: detection and characterization by structure function analysis 1708.08703

Maroussia Roelens¹, Laurent Eyer¹, Nami Mowlavi¹, Isabelle Lecoeur-Taïbi², Lorenzo Rimoldini², Sergi Blanco-Cuaresma¹, Lovro Palaversa³, Maria Süveges⁴, Jonathan Charnas², Thomas Wevers⁵

¹Observatoire de Genève, Département d'Astronomie, Université de Genève, Chemin des Maillettes 51, CH-1290 Versoix, Switzerland

h between them. The light-curve consists in magnitudes $(m_i)_{i=1..n}$ observed at times $(t_i)_{i=1..n}$. The variogram value for a time lag *h* is denoted by $\gamma(h)$. It is defined as the average of the squared difference in magnitude $(m_j - m_i)^2$ computed on all pairs (i, j) such that $|t_j - t_i| = h \pm \epsilon_h$ where ϵ_h is the tolerance accepted for grouping the pairs by time lag. This binning can be necessary, particularly in the case of uneven time sampling, to make sure to have enough pairs to compute variance for a given lag. If we note N_h the number of such pairs,

$$\gamma(h) = \sum_{i>j} \frac{(m_j - m_i)^2}{N_h} \tag{1}$$

Variogram

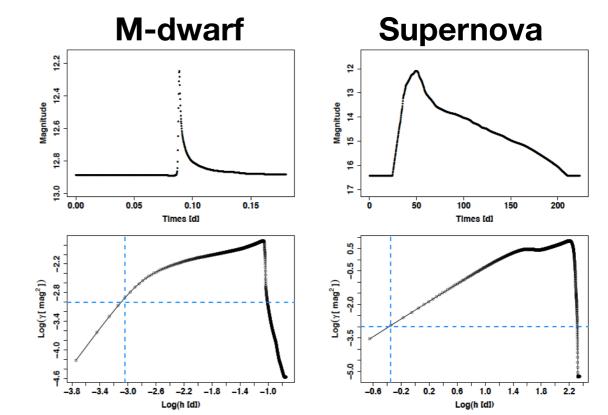
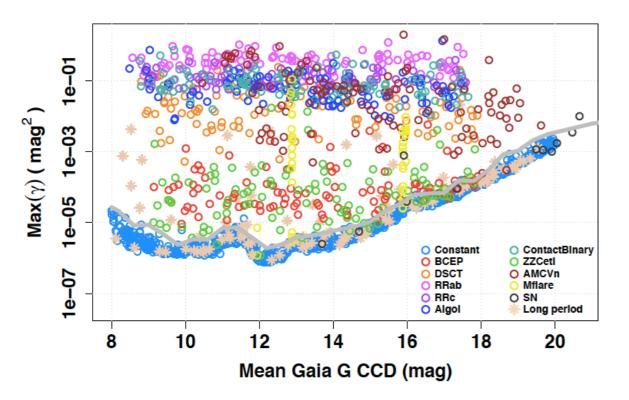


Figure 6. Example of M flare continuous light-curve (top), and corresponding theoretical variogram (bottom). The blue dashed lines indicate the detection threshold ($\gamma_{det} = 10^{-3} \text{ mag}^2$) and the associated detection timescale τ_{det} .

Figure 7. Example of SN continuous light-curve (top), and corresponding theoretical variogram (bottom). The blue dashed lines indicate the detection threshold ($\gamma_{det} = 10^{-3} \text{ mag}^2$) and the associated detection timescale τ_{det} .

x-axis extents differ



Short timescale variability can be recovered

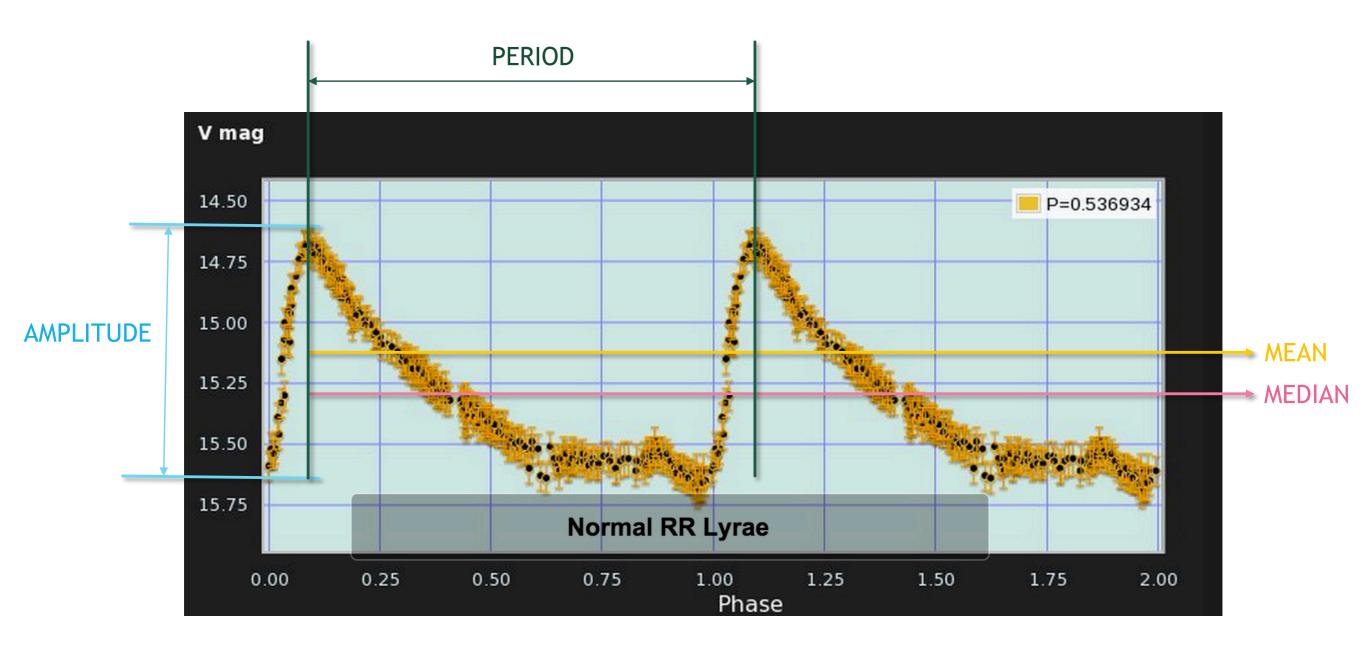
Figure 17. Maximum weighted variogram value as function of the mean magnitude, for the short timescale periodic variables, the transients, the longer period variables and the constant sources of the *Gaia*-like data set. The grey line shows the refined detection threshold depending on the mean magnitude of the source ($\gamma_{det} = \Gamma(\bar{G}_{CCD})$).

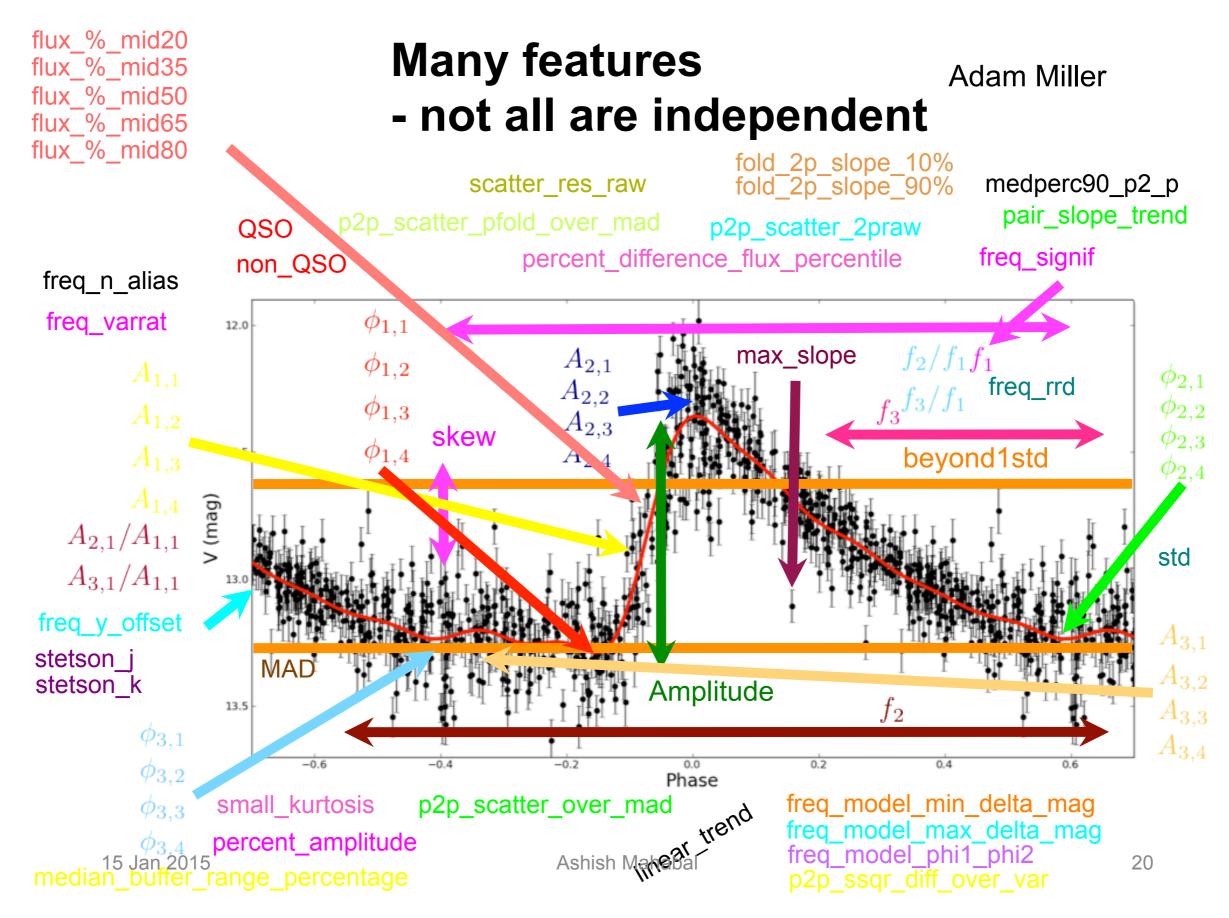
Contamination from longer term variability (removable through postprocessing)

6 CONCLUSION

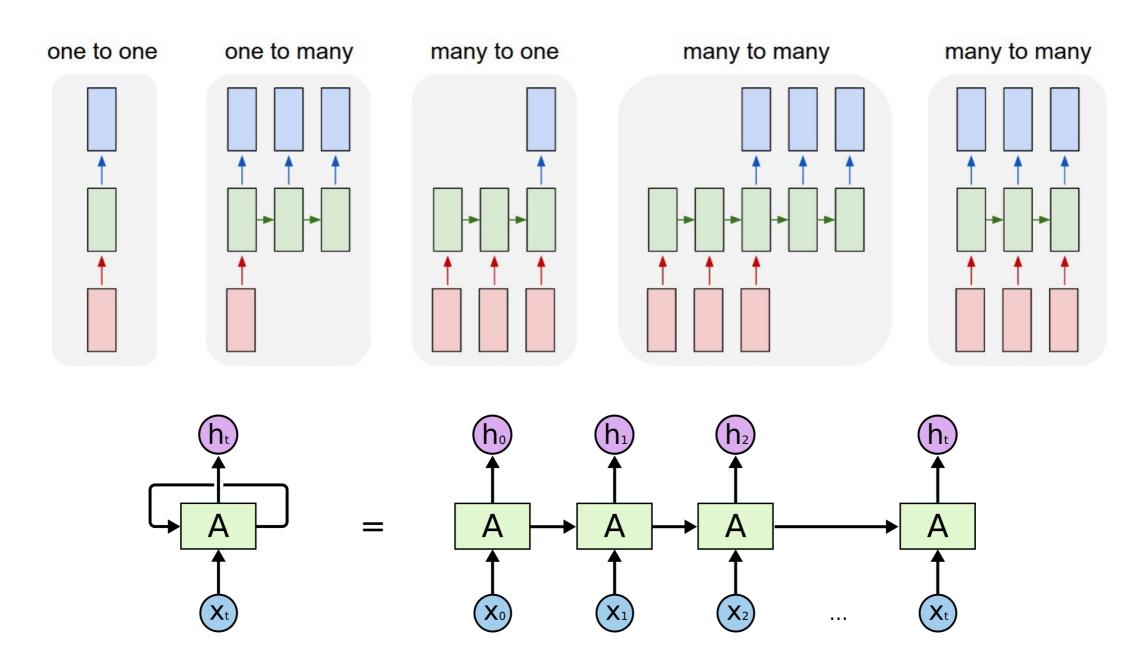
In this work, by mean of extensive light-curve simulations, we showed that, with a specifically tailored detection criterion, the variogram method should enable a good recovery of short timescale variability, periodic or transient, from *Gaia* per-CCD photometry, with a reduced fraction of observed constant sources resulting in false positives. Contamination for longer period variables is significant, and is essentially due to amplitude variations greater than 0.25 mag typically. It should be efficiently eliminated by post-processing involving both periodogram investigation and comparison with the other variability studies performed in the *Gaia* DPAC con-

Light-curve features

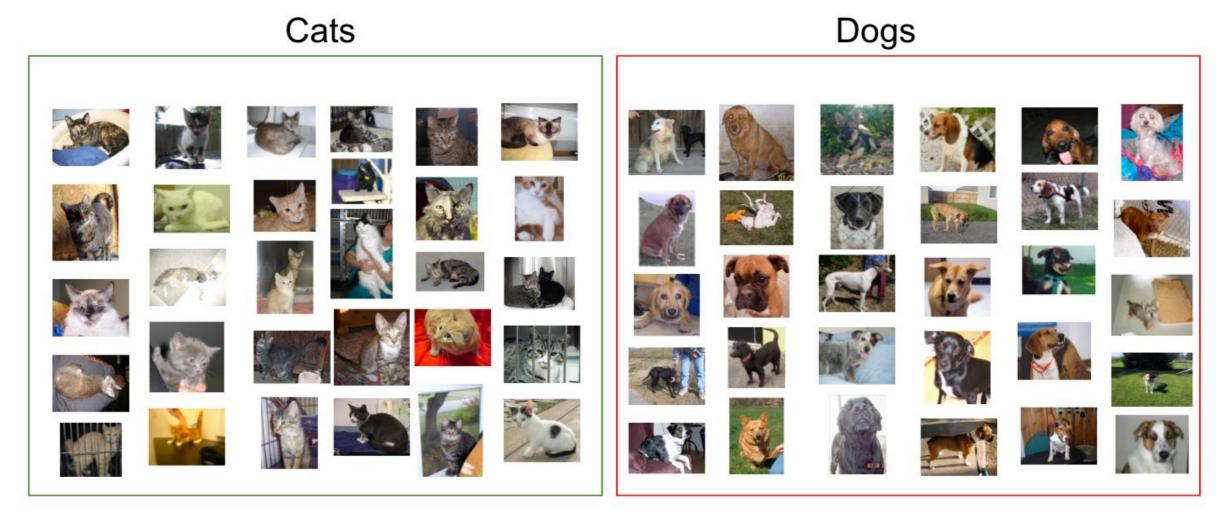




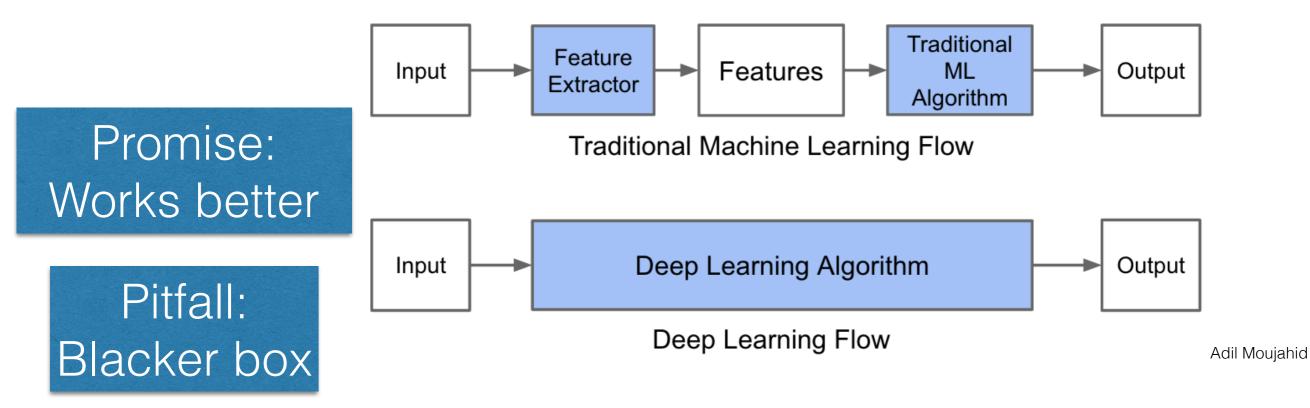
Recurrent Neural Networks



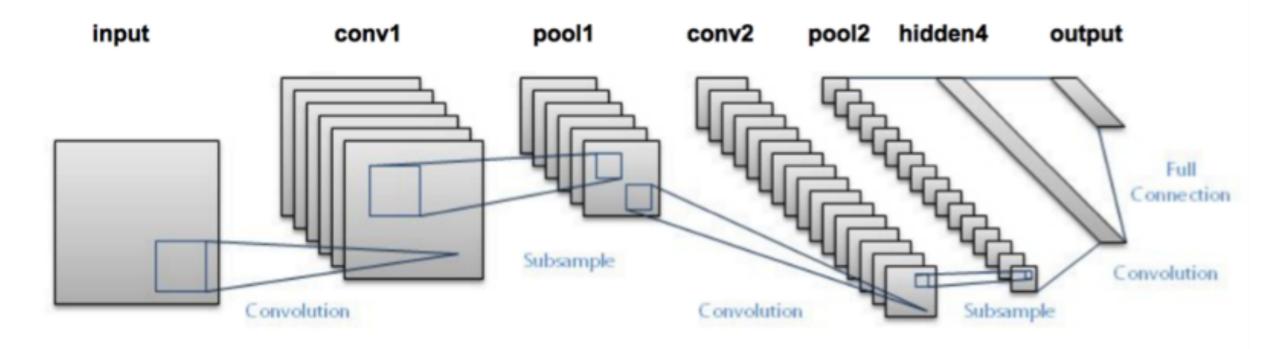
http://colah.github.io/posts/2015-08-Understanding-LSTMs/



Sample of cats & dogs images from Kaggle Dataset



Convolutional network (single slide) primer



analyticsvidhya.com

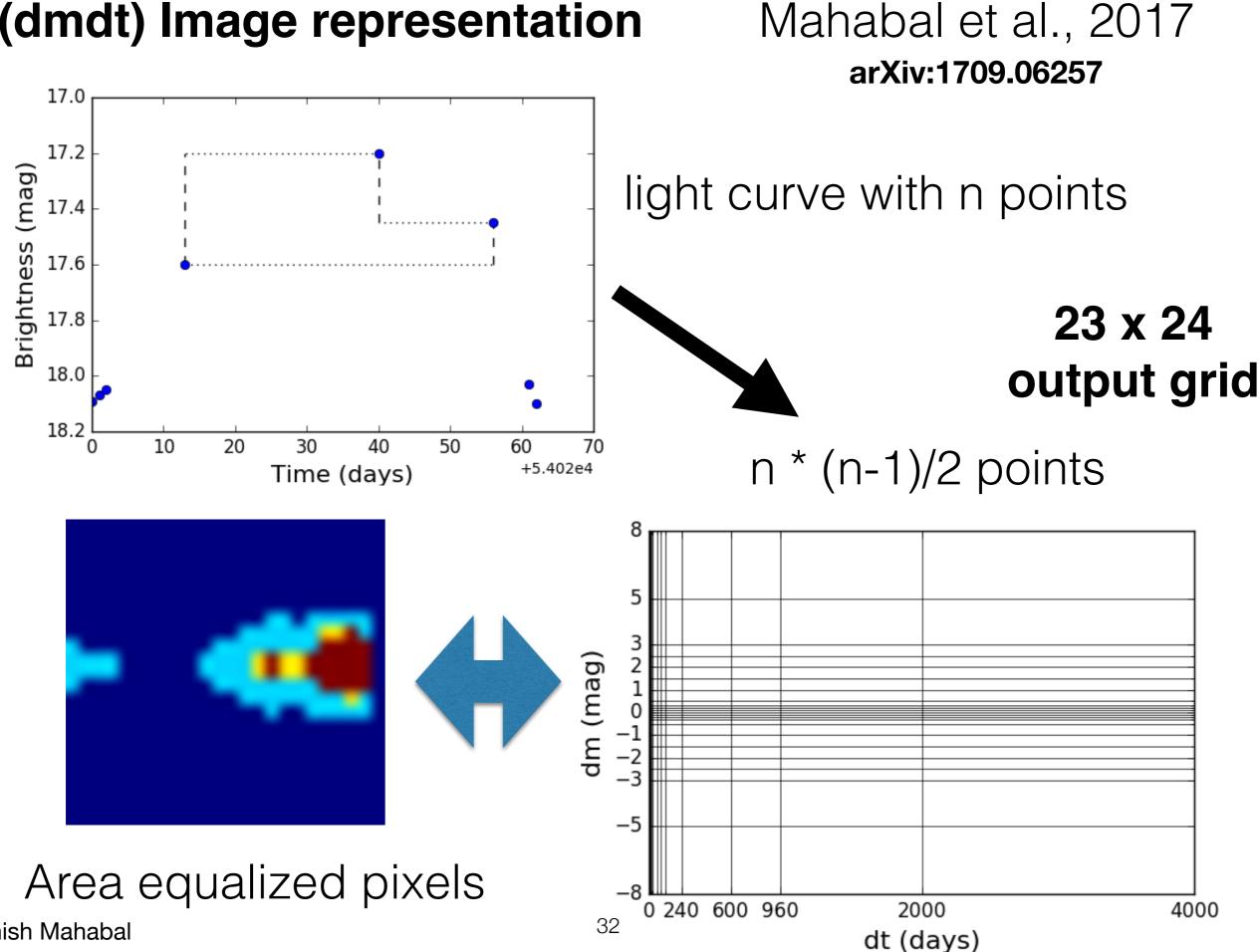
INPUT IMAGE								
18	54	51	239	244	188			
55	121	75	78	95	88			
35	24	204	113	109	221			
3	154	104	235	25	130			
15	253	225	159	78	233			
68	85	180	214	245	0			



429

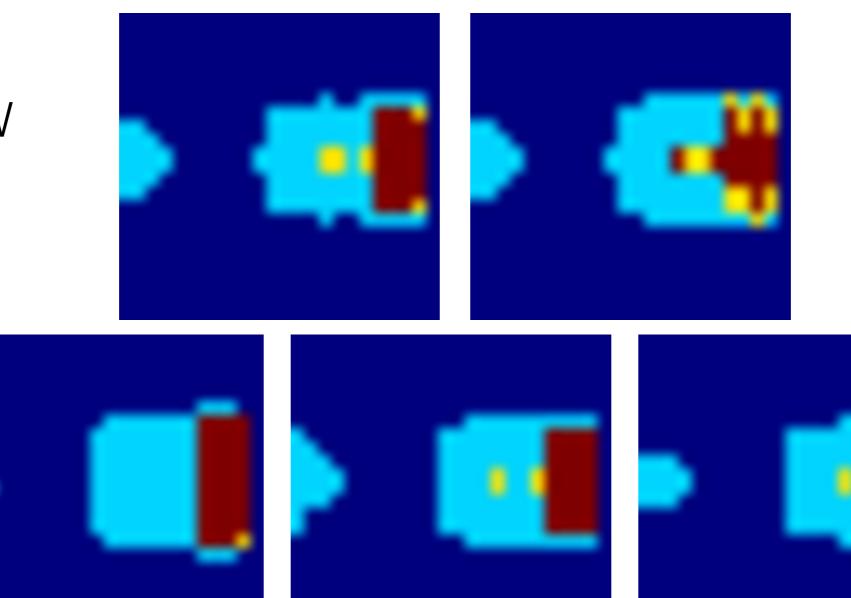
429	505	686	856
261	792	412	640
633	653	851	751
608	913	713	657

792	856
913	851



(dmdt) Image representation

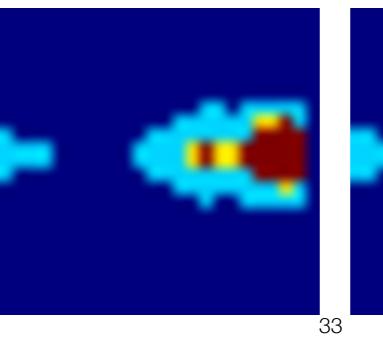


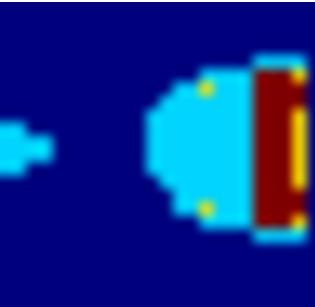




RS CVn

with Kshiteej Sheth Ashish Mahabal

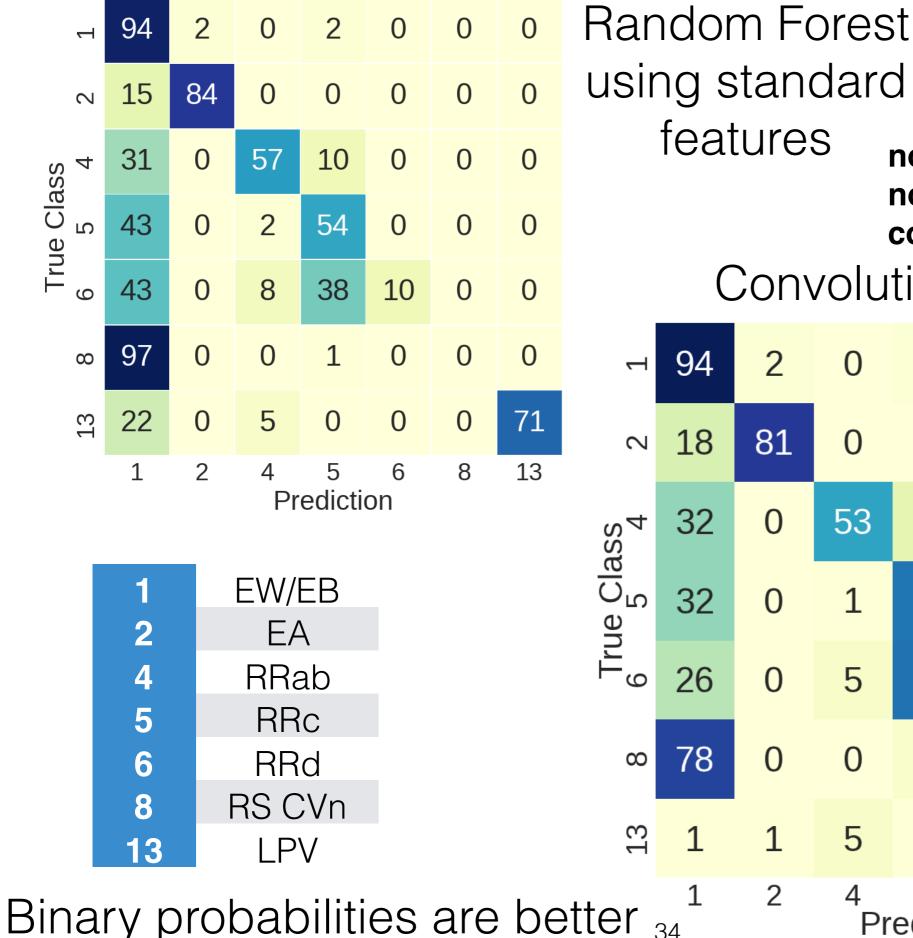




LPV

ΕA

medians



no features no dimensionality reduction comparable results **Convolutional Network**

1	94	2	0	2	0	0	0
2	18	81	0	0	0	0	0
ISS 4	32	0	53	14	0	0	0
True Class 6 5 4	32	0	1	65	0	0	0
Tru 6	26	0	5	66	0	1	0
8	78	0	0	4	0	13	0
13	1	1	5	1	2	3	83
ter	1 34	2	4 Pr	5 edictio	6 on	8	13

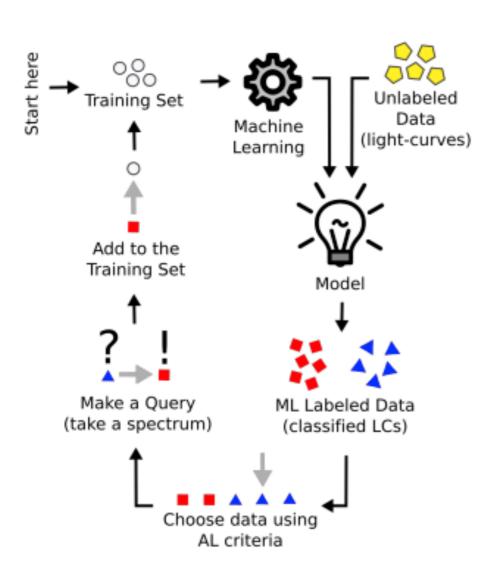
Visualizations (astronomy)

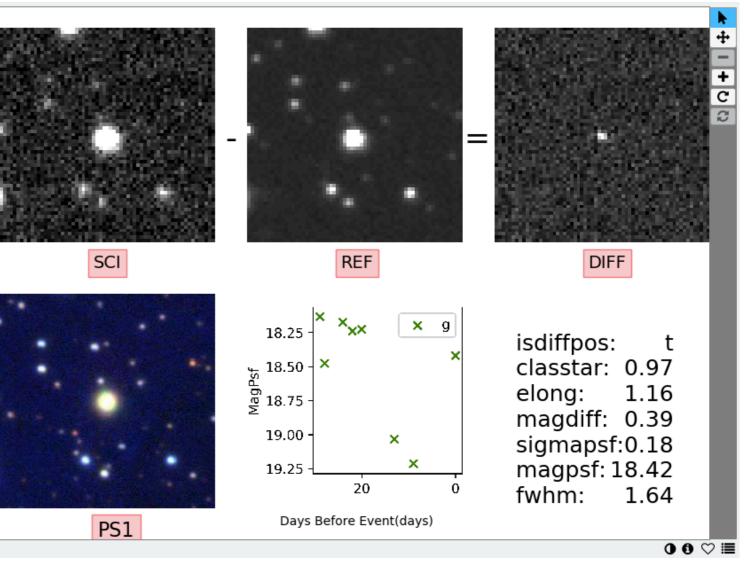
Interpretability

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
°	• •		6 9 6	•	•	•				•				• •••	15
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
9 	•	•	•	°	°		°		•					•	31 •
32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
°	9 9 900		•		0 9 9 9	•				9 			***	0 9 	47
48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
6	0 - 0 - 1	0 	1	°		9 9 60		0	0 	°	•	° °	9 200	•	63 •••
64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
6 9 200		•	100 - 100 -			0 0 124	• •	• •			°	9			79
80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95
		9 - 100	•		•	•	•	•	°	•	•	•		•	() 200

TransiNet with N Sedaghat

Active learning to minimize follow-up





Images, light curves, metadata

Ishida et al. 2018, arXiv:1804.03765

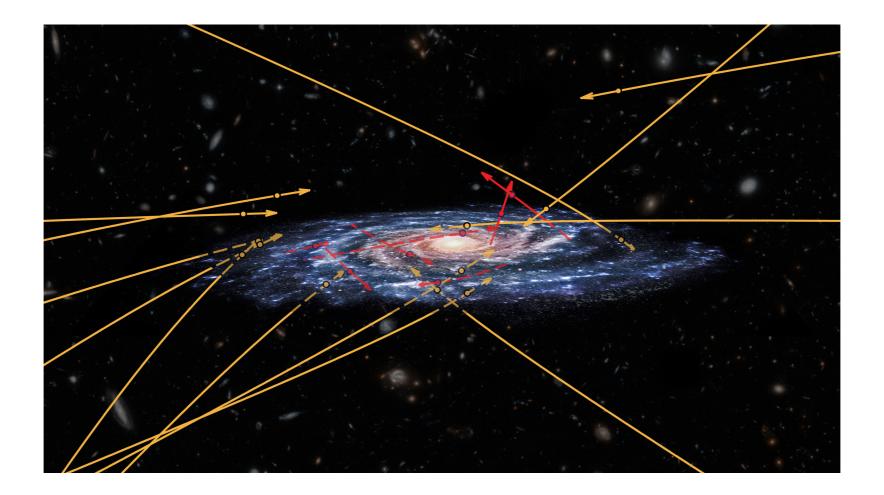
Accuracy of 80% reached in 100 days of observations, far above the canonical rate

ed in s, far ate Connecting to Brokers ALeRCE.science AMPEL Antares

36

Antares Lasair LSST data challenges (check on Kaggle to win 25K USD) 1810.00001

Hypervelocity stars and high cadence data

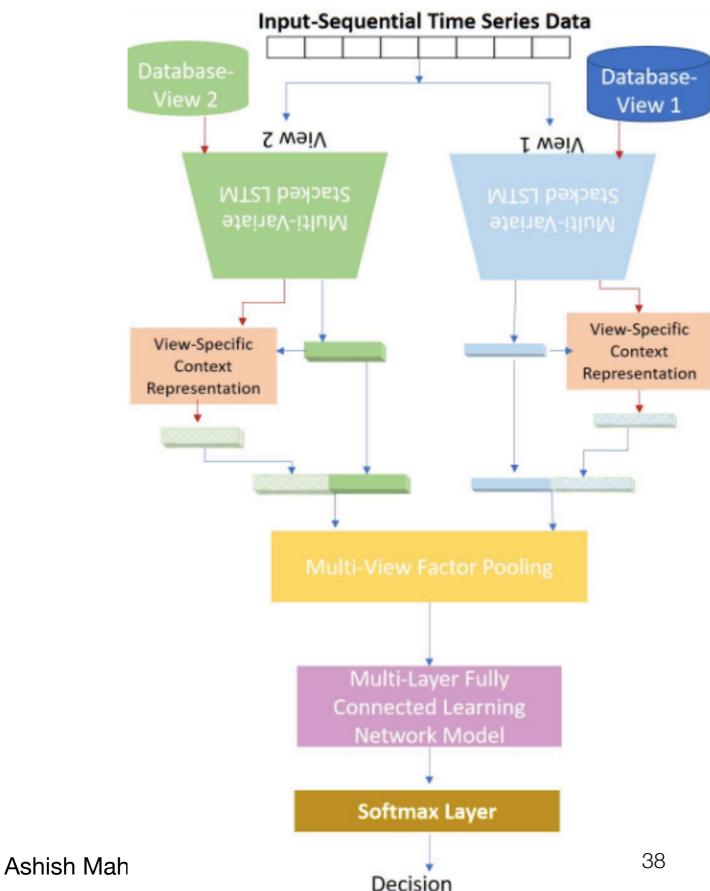


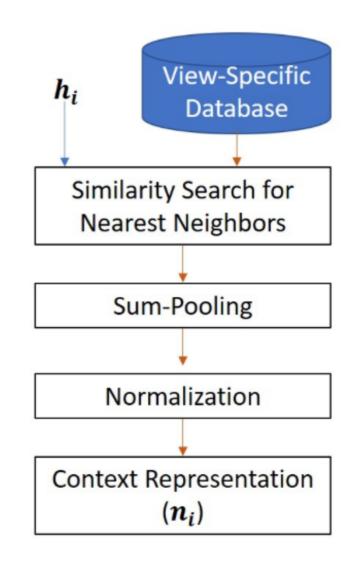
Tomasso Marchetti's talk last week (Caltech)

Possibly extra-Galactic stars

Many more epochs could help here as well (over a much longer time)

VIFI and CRTS/PTF/Gaia/ZTF





Multi-view transfer learning

S Das et al. 2018

White dwarf sentinel service

ZTF White Dwarfs Home Search



White Dwarfs detected with the Zwicky Transient Facility

This website provides access to ZTF alerts, cross-matched with a Gaia DR2-based catalog of white dwarfs, on a nightly basis.

+	 ● September 9, 2018 - October 8, 2018 - Search 											- 12 +	
	T_obs, UTC 💧	objectId 🖕	candid 🔶	name 🔶	pwd	Gaia mag 🖕	candidate						
	1_005, 010						fid	rb≑	magpsf 🖕	sigmapsf 🖕	R.A.	Decl.	
+	2018/09/09_03:17:19	ZTF18aaavxnm	616137034215015006	WDJ150241.00+333424.00	0.31	17.34159	2	0.79	18.99	0.14	15:02:40.8423	33:34:23.596	
+	2018/09/09_03:34:21	ZTF18abujqpr	616148865015015018	WDJ185455.06+212519.96	0.99	16.92406	2	0.68	19.20	0.14	18:54:54.9461	21:25:20.488	
+	2018/09/09_03:35:00	ZTF18abguzyw	616149312515010001	WDJ182830.24+260125.33	0.06	18.72079	2	0.69	20.10	0.15	18:28:30.2412	26:01:25.384	
+	2018/09/09_03:50:03	ZTF18aamigoa	616159762015010012	WDJ181824.35+465721.46	0.75	17.50169	2	0.64	20.05	0.18	18:18:24.3484	46:57:21.709	
+	2018/09/09_03:53:23	ZTF18aaiytds	616162080515010008	WDJ172758.14+380022.43	0.07	18.04158	2	0.77	16.45	0.05	17:27:58.1241	38:00:22.585	
+	2018/09/09_03:54:05	ZTF18aaimyrh	616162554515010003	WDJ174033.46+414755.98	0.56	19.67580	2	0.69	19.99	0.19	17:40:33.4570	41:47:55.891	
+	2018/09/09_04:03:24	ZTF18abbwsdx	616169024515010012	WDJ181821.02+125350.27	0.45	19.05067	1	0.63	20.18	0.17	18:18:21.0462	12:53:50.403	
+	2018/09/09_04:52:32	ZTF18aabtvvb	616203151915015002	WDJ183304.07+463705.37	0.24	15.75325	1	0.43	17.81	0.09	18:33:04.1277	46:37:04.673	
+	2018/09/09_04:52:32	ZTF18aabtvvb	616203151915015003	WDJ183304.07+463705.37	0.24	15.75325	1	0.51	17.78	0.04	18:33:04.1159	46:37:04.994	
+	2018/09/09_05:20:13	ZTF18aabtvvb	616222373015015006	WDJ183304.07+463705.37	0.24	15.75325	1	0.71	17.03	0.06	18:33:04.0855	46:37:05.043	
+	2018/09/09_05:31:49	ZTF18abukmkl	616230436215010005	WDJ221508.00-073148.54	0.90	19.60407	1	0.75	20.03	0.19	22:15:07.9988	-7:31:48.541	

ZTF alerts, matched to Gaia DR2

https://rico.caltech.edu



Take away message

- There are areas where working together will benefit Gaia and ZTF science
- In particular to check iffy and/or strange transients in the other survey
- Combining ZTF cadence and Gaia's multi-epoch positional accuracy can be rewardi
- I am sure I have left out many things (that I do not know enough about)
- Lets talk ...

aam at astro.caltech.edu

