



# ALeRCE

Automatic Learning for the  
Rapid Classification of Events

## The universe in a stream: challenges & progress of the ALeRCE broker.

F. Förster; G. Cabrera-Vives; P. A. Estévez; P. Sánchez-Sáez; J. Arredondo; F. E. Bauer; R. Carrasco-Davis; M. Catelan; F. Elorrieta; S. Eyheramendy; P. Huijse; A. Moya; G. Pignata; E. Reyes; I. Reyes; D. Rodríguez-Mancini; D. Ruz-Mieres; C. Valenzuela; I. Álvarez-Maldonado; N. Astorga; A. Bayo; J. Borissova; F. Chabour; A. Clocchiatti; M. Contreras; R. Dastidar; D. De Cicco; C. Donoso-Oliva; M. J. Graham; L. Hernández-García; R. Kurtev; A. Mahabal; R. Molina-Ferreiro; A. Muñoz-Arancibia; W. Palma; M. Pérez-Carrasco; A. Papageorgiou; J. Pineda; P. Protopapas; M. Romero; L. Sabatini-Gacitua; A. Sánchez; J. San Martín; K. Sharma; J. Silva; E. Vera; J. R. Vergara; + C. Ashall; A. Belinski; Th. de Jaeger; G. Folatelli; Ll. Galbany; S. González-Gaitán; N. Ikonnikova; A. Mourao; B. Safonov; N. Shatsky; A. Tatarnikov; D. Tsvetkov; O. Vozyakova; S. Zheltouhov

<http://alerce.science/>

**CMM**  
Center for  
Mathematical  
Modeling



**DO**  
DATA OBSERVATORY

**U**  
LISBOA  
UNIVERSIDADE  
DE LISBOA



**ICE** INSTITUT DE  
CIÈNCIES  
DE L'ESPAI  
**CSIC** **IEEC**



UNIVERSIDAD  
NACIONAL  
DE LA PLATA



**ROUNA**  
Ciencia y Educación en Red



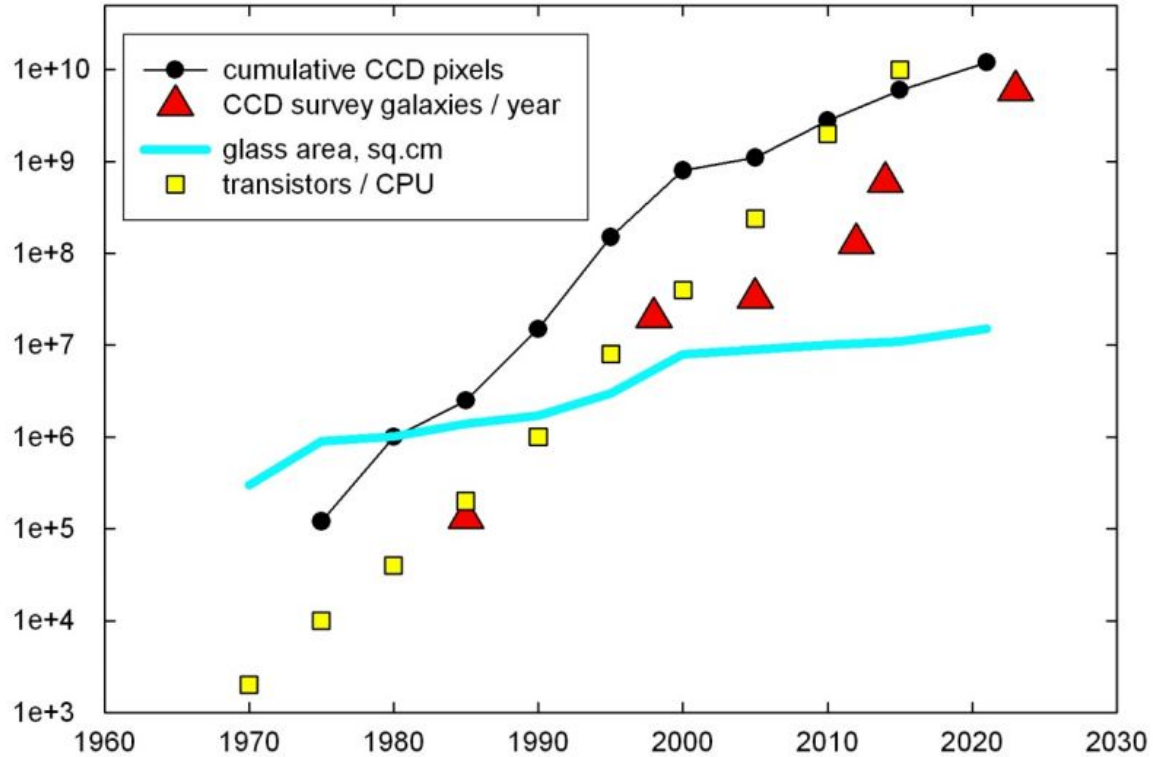
**NLHPC**  
National Laboratory  
for High Performance  
Computing  
Chile







# Moore's law vs mirrors

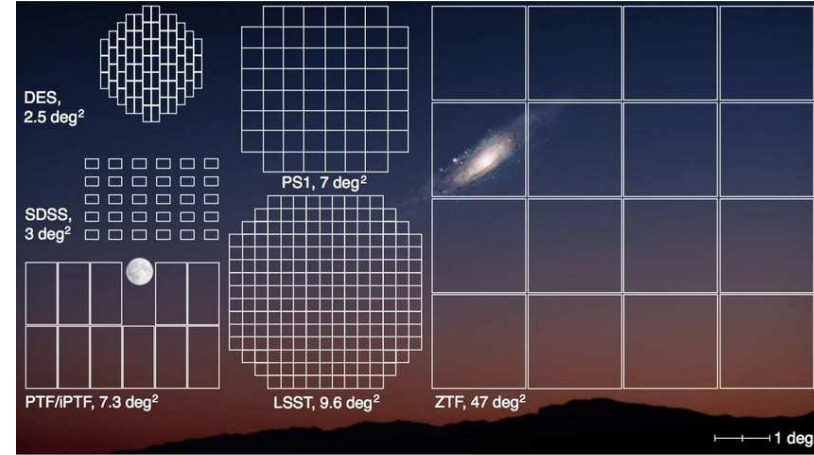
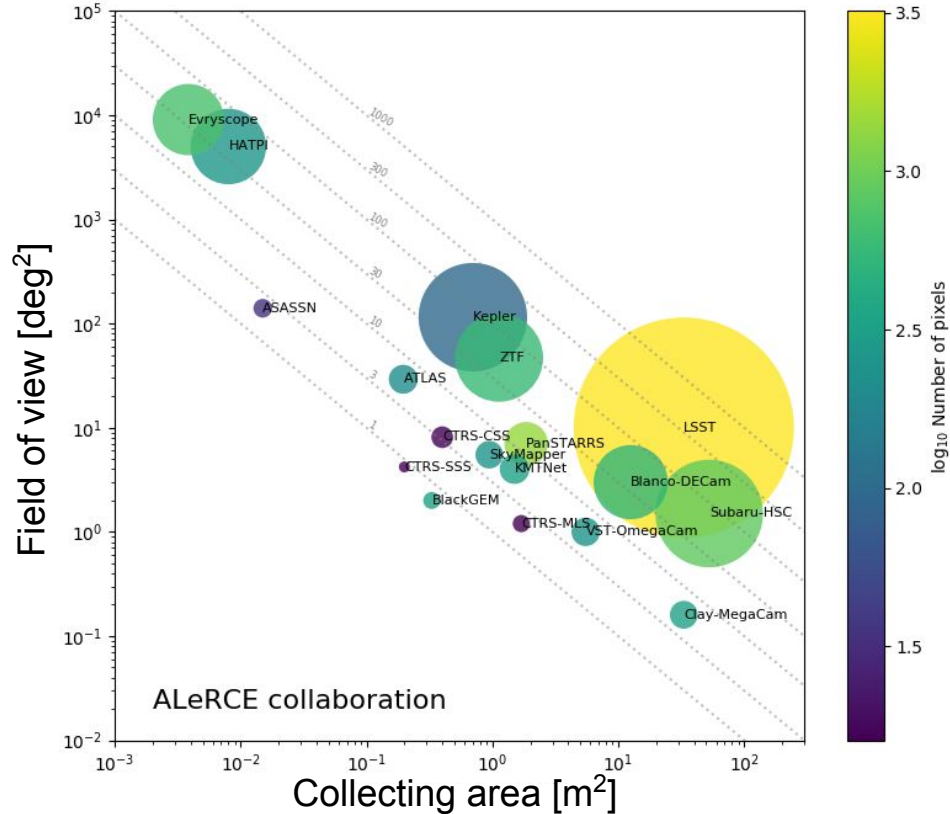


Tyson et al. 2019

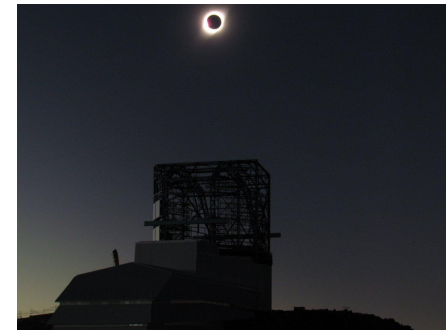


# Optical survey telescopes

Circle size: etendue

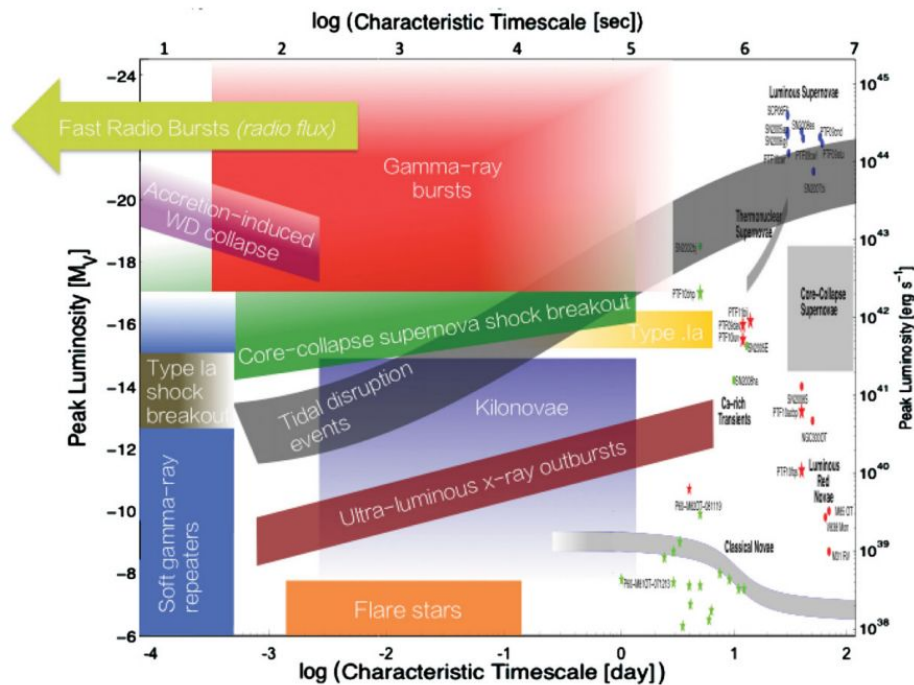
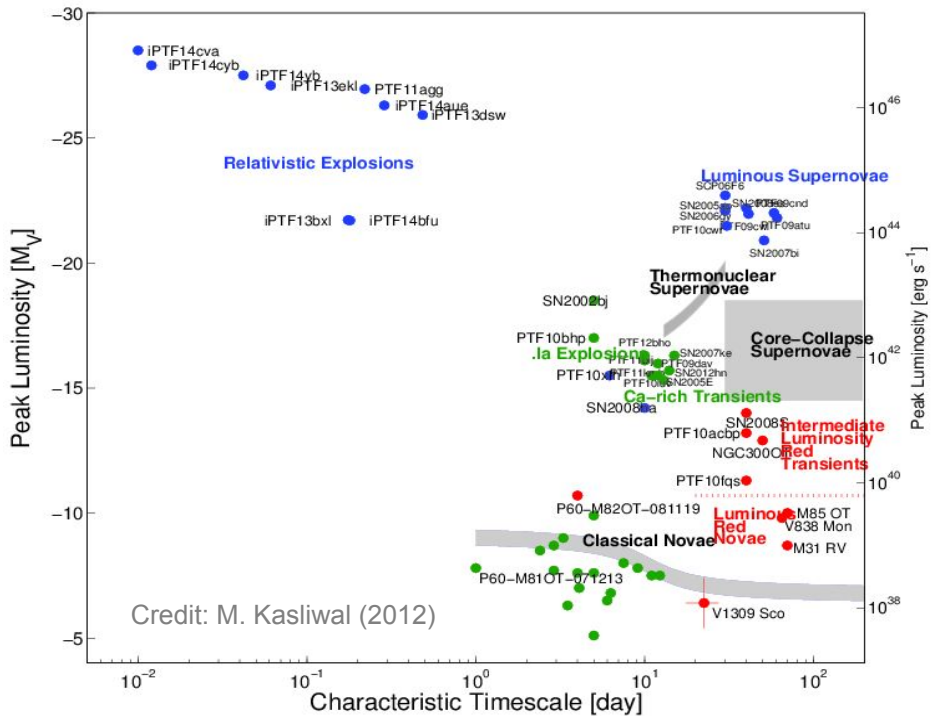


Vera C. Rubin Observatory





# Transient landscape



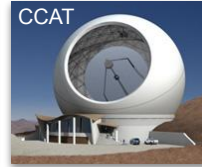
Burton+2016



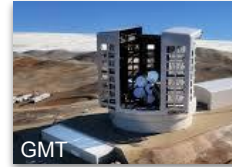
# Astronomical infrastructure in Chile



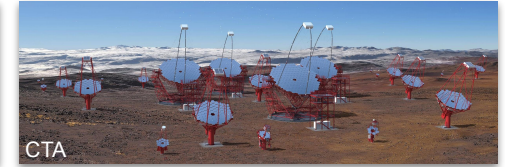
Magellan



CCAT



GMT



CTA

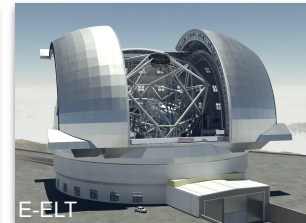


VISTA

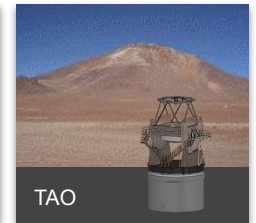
VLT



VRO



E-ELT



TAO



SOAR

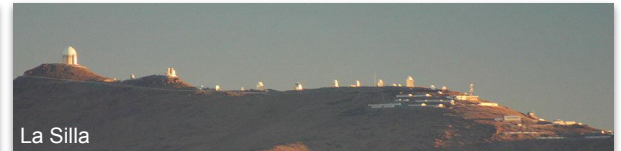
Gemini



ALMA



CTIO



La Silla

Chilean institutions: access to ~10% observing time



# Future time domain astronomy ecosystem

Survey telescopes



Alert brokers/TOMs



+ *other brokers/TOMs*



Follow up telescopes





# Tools for time domain astronomy

Acquisition & processing



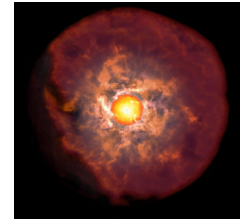
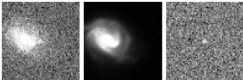
Alert filtering & classification



Prioritization & follow-up



Physical interpretation



**Survey telescopes:**  
image processing,  
real/bogus filtering

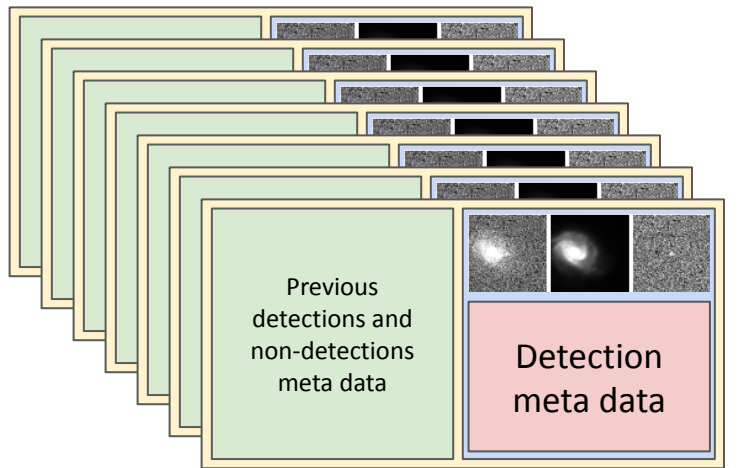
**Brokers:**  
aggregation,  
crossmatching,  
ML classification

**TOMs & follow-up telescopes:**  
resource optimization  
& communication  
(APIs), actionable ML

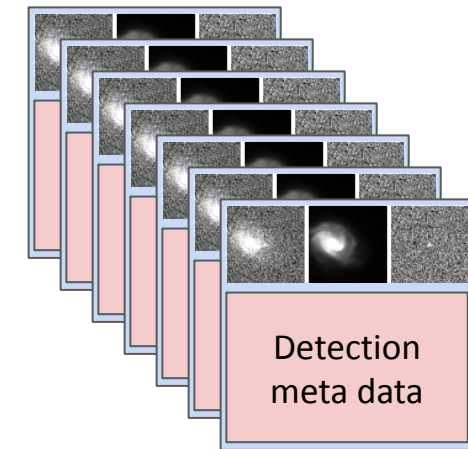
**Analysis:**  
modeling, inference  
(e.g. MCMC),  
prediction



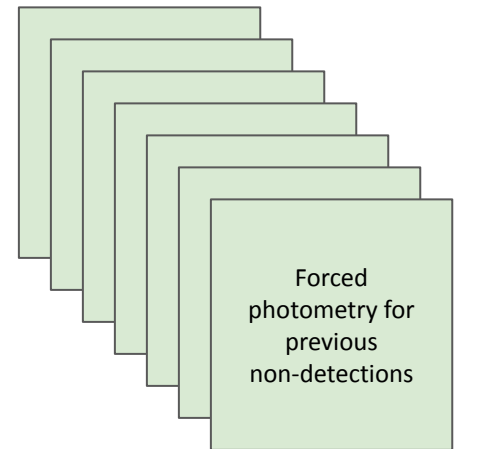
# Alerts



**ZTF stream:**  
~300 k per night  
~15 minutes from observation



**LSST stream:**  
~10 M per night  
60 seconds from observation

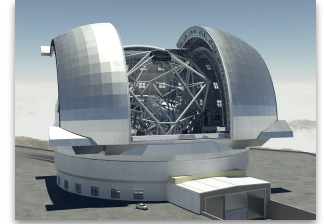


**LSST Science Platform:**  
24 hours from observation





# Interoperable tools for new discoveries





Pitt-Google


LasAIR



ALeRCE  
Automatic Learning for the  
Rapid Classification of Events

**Vera C. Rubin officially selected Community Brokers**

<https://www.lsst.org/scientists/alert-brokers>

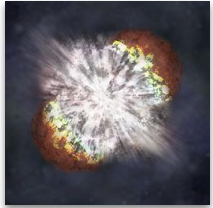


ALeRCE is a Chilean-led broker officially selected as one of the Community Brokers for the Vera Rubin Observatory and its Legacy Survey of Space and Time (LSST).

Presentation paper: Förster et al. 2021, AJ; Stamp classifier: Carrasco-Davis et al. 2021, AJ; Light curve classifier: Sánchez-Sáez et al. 2021, AJ



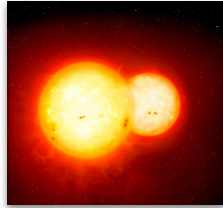
# Scientific questions



## Transients

Progenitors of stellar explosions (outermost layers) & explosion physics (ejecta structure).

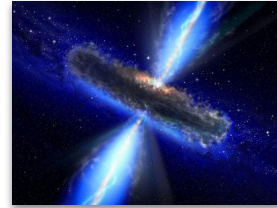
**How does the Universe expand?**



## Variable stars

Microlensing events, changing mode stellar pulsators, rapid reaction to eclipsing events, eruptive events.

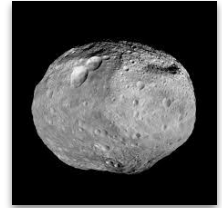
**How is matter distributed in our galaxy?**



## Supermassive black holes

Changing state AGNs, reverberation mapping, detection of intermediate mass black holes, tidal disruption events.

**How do black holes affect the evolution of galaxies?**



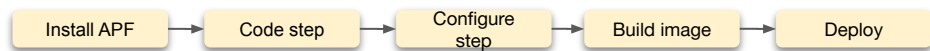
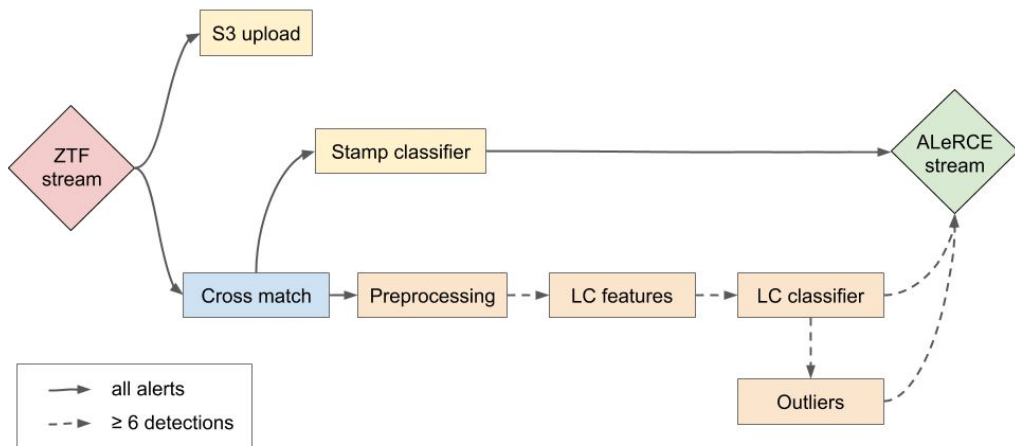
## Asteroids

Near earth objects, size distribution, collisional or rotational disintegration.

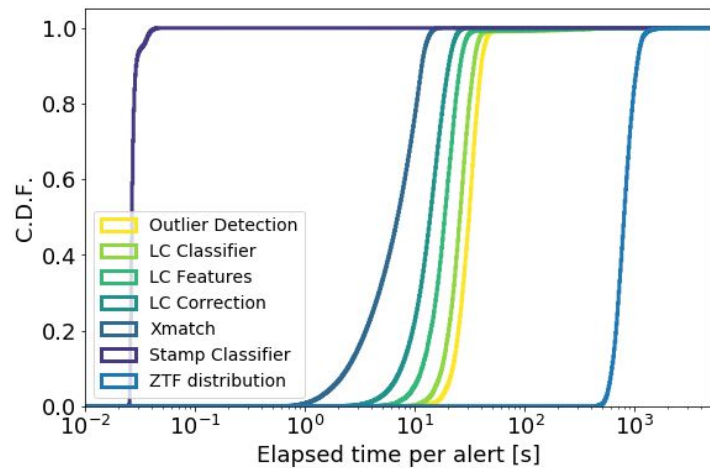
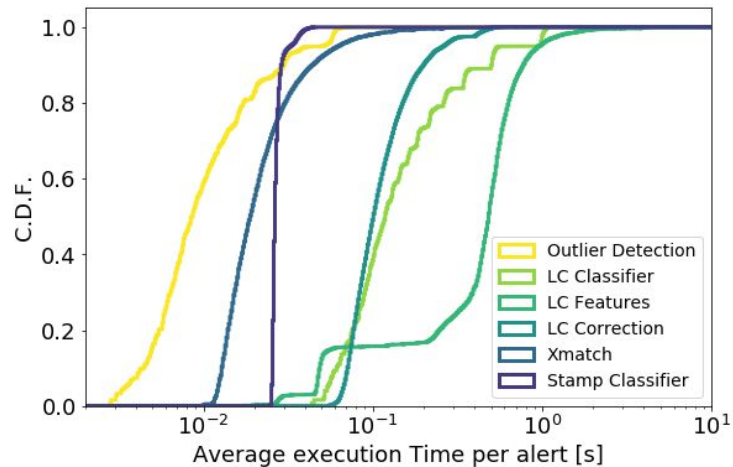
**How was the solar system formed?**



# ALeRCE pipeline



<https://apf.readthedocs.io/en/latest/>



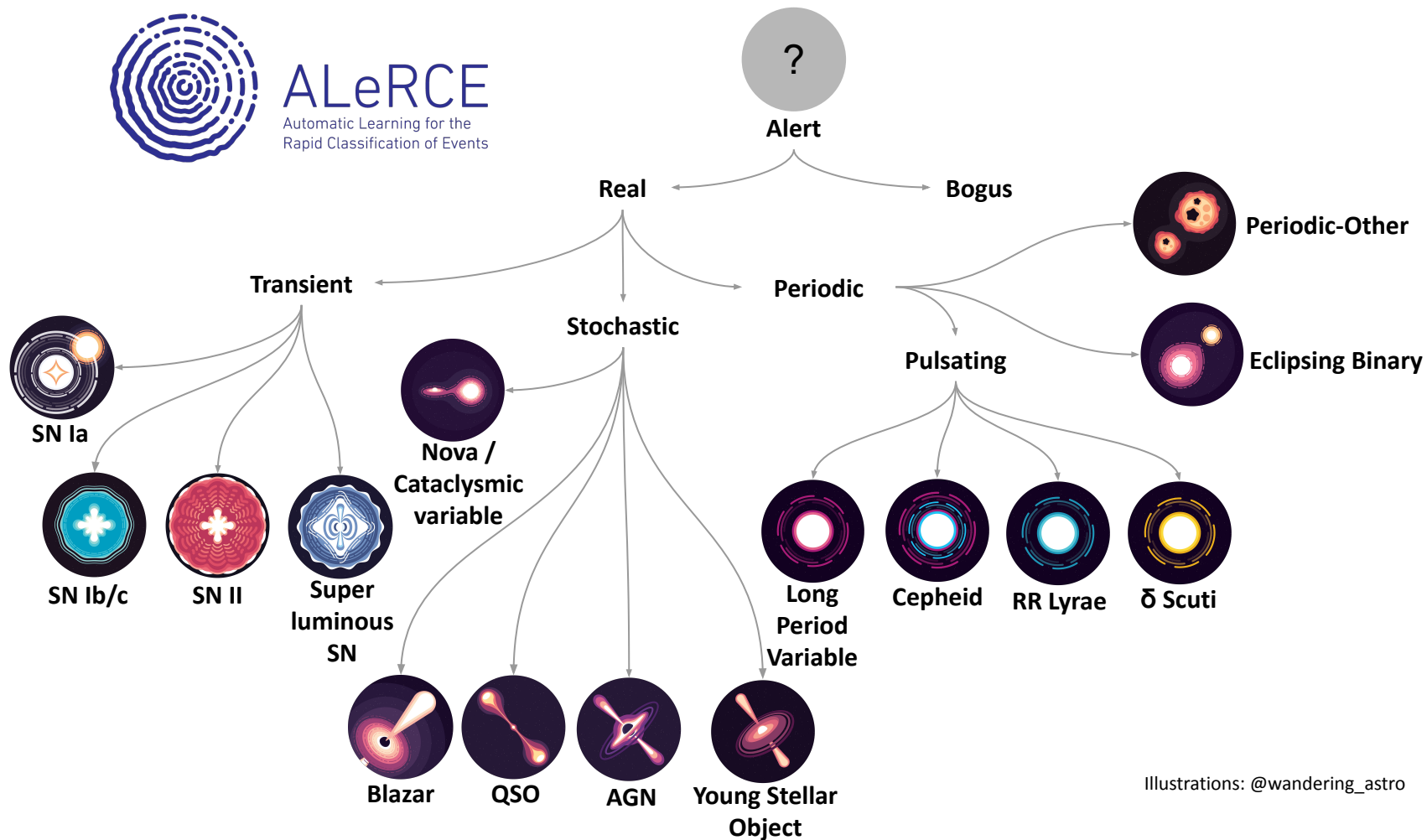






# ALeRCE

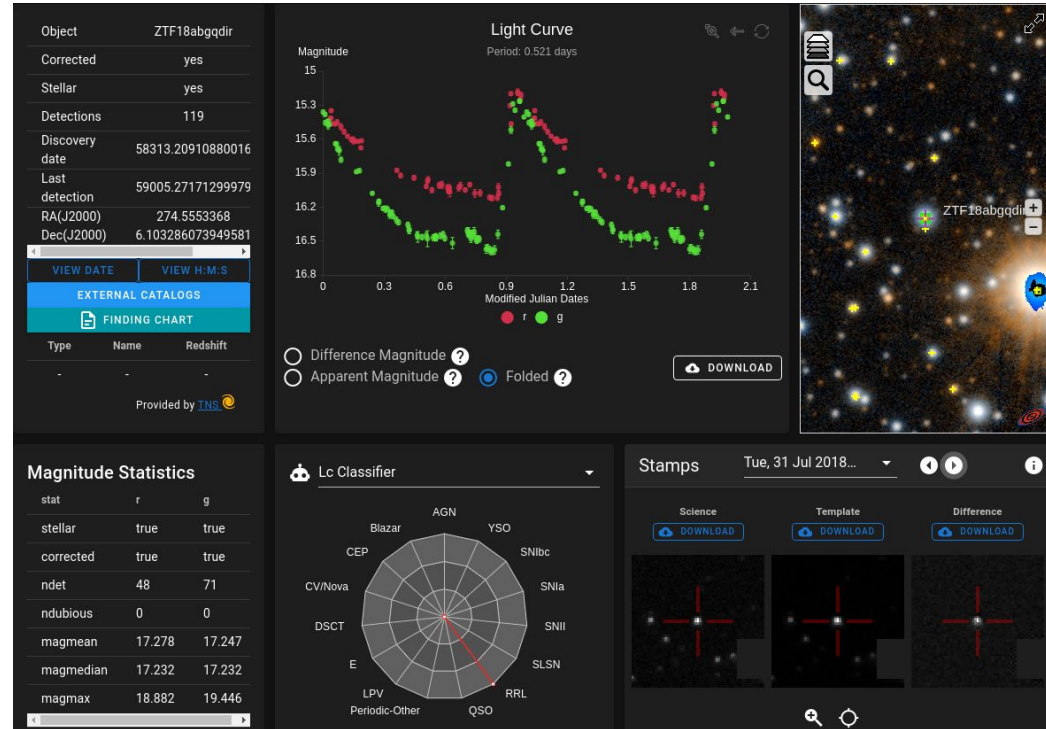
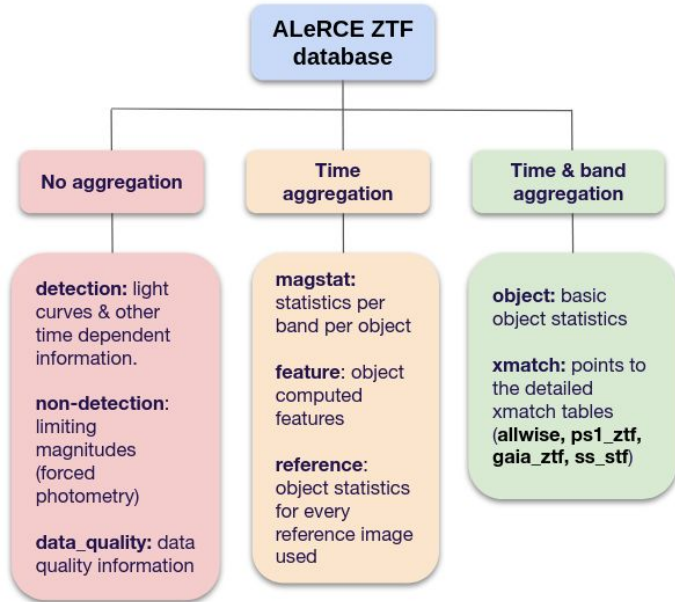
Automatic Learning for the  
Rapid Classification of Events



Illustrations: @wandering\_astro



# Database, API, client, & explorer



**Database:** PostgreSQL + other noSQL

**API:** <http://api.alerce.online/ztf/v1>

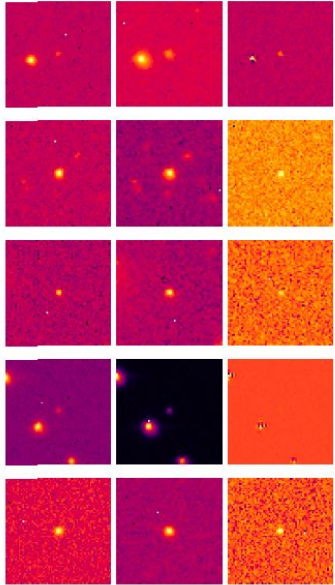
**Python client:** <https://alerce.readthedocs.io/en/latest/>

**Explorer (powered by API):** <http://alerce.online>

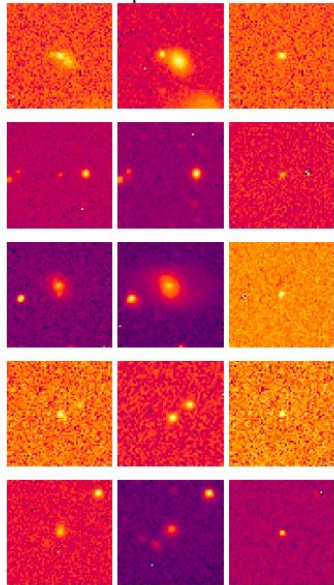


# Image stamps

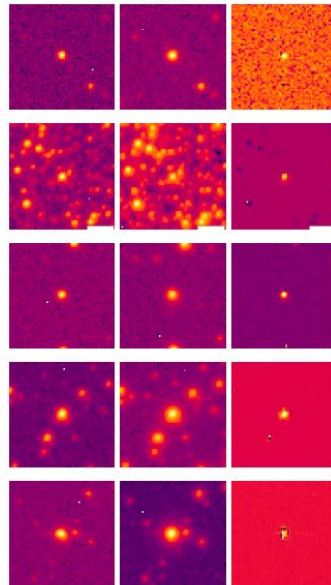
Active Galactic Nuclei



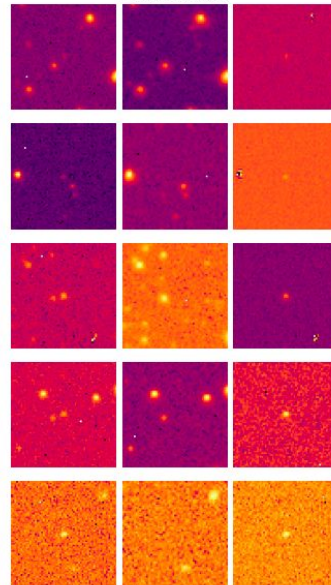
Supernovae



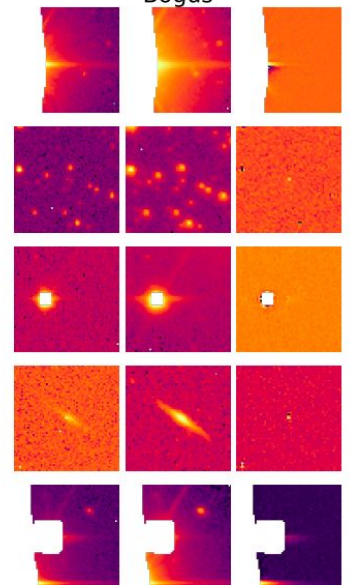
Variable Stars



Asteroids



Bogus

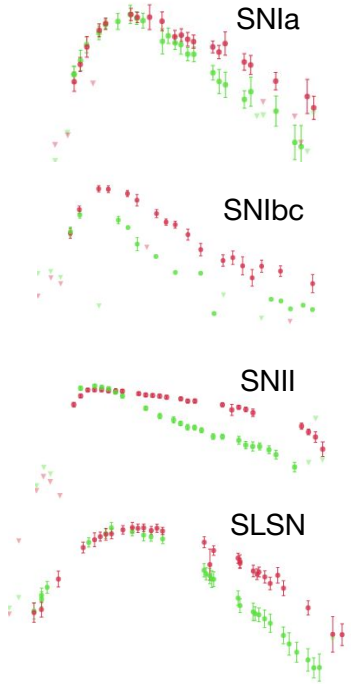




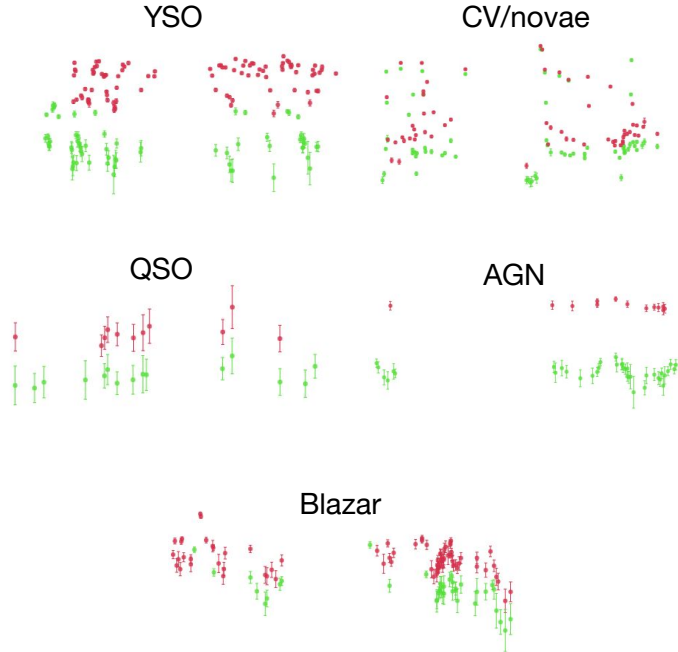
# Light Curves



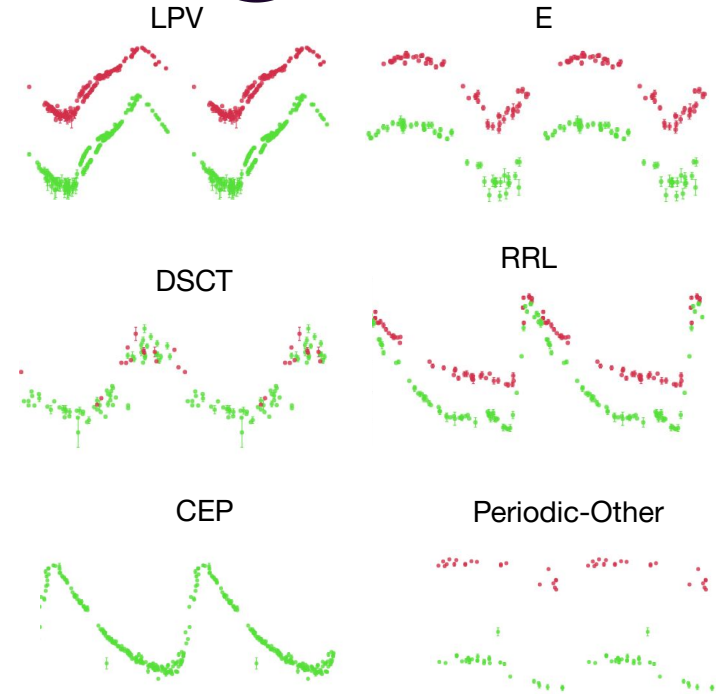
**Transient**



**Stochastic**



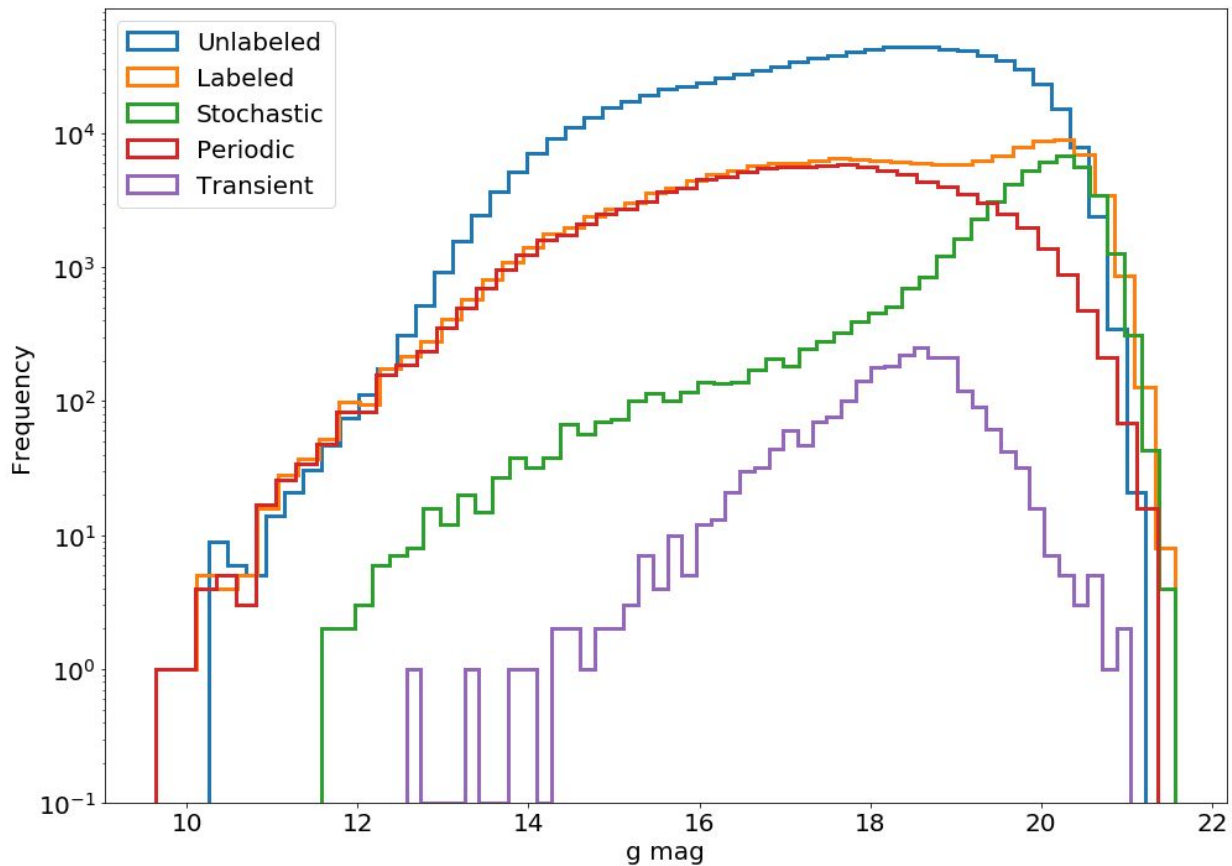
**Periodic**  
(folded LCs)







# Training set: magnitude distribution



LSST  
25th mag!

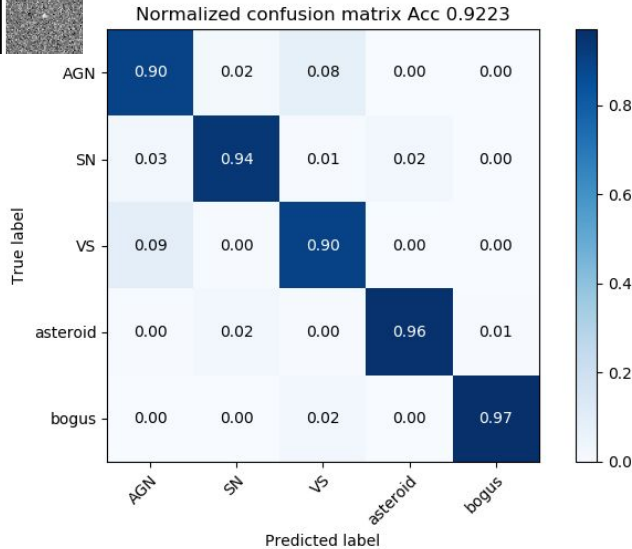
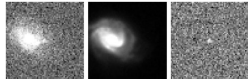




# Stamp based classifier

*Convolutional neural network*  
Carrasco-Davis et al. 2020

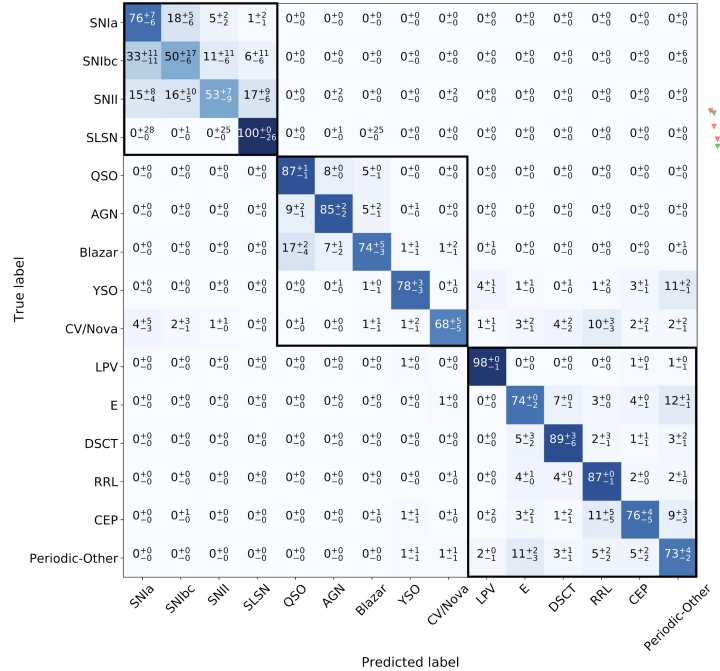
AGN, SN, VS, asteroid, bogus



# Light curve based classifier

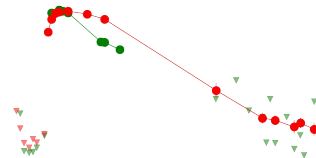
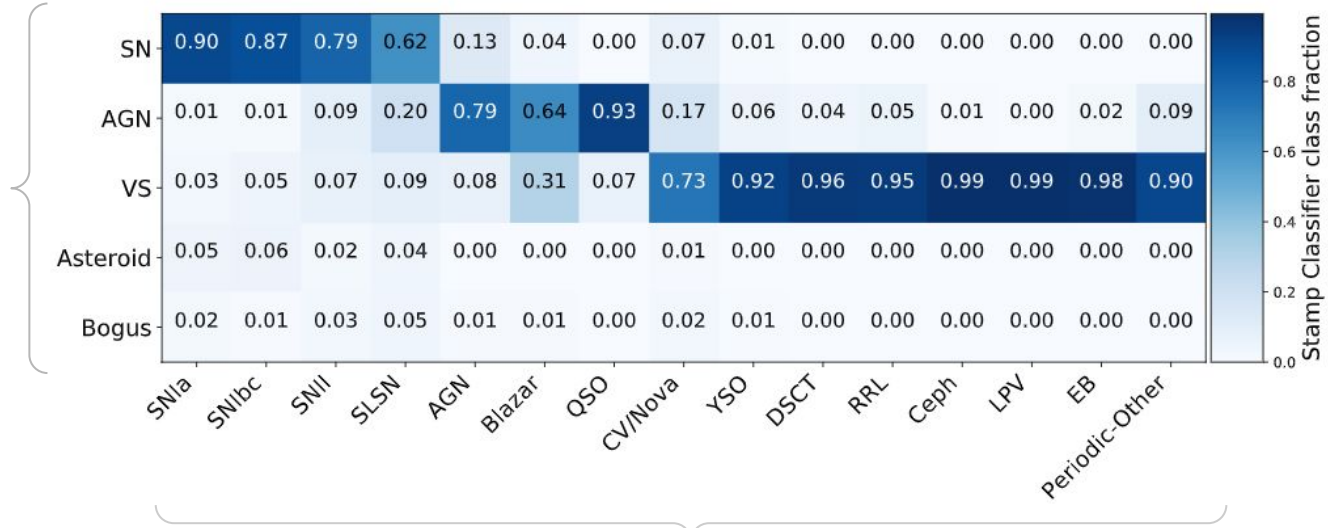
*Balanced Hierarchical Random Forest*  
Sánchez-Sáez et al. 2020

SN Ia, SN Ibc, SN II, SLSNe, QSO, AGN, Blazar, YSO, CV/Nova, LPV, E, DSCT, RRL, CEP, Periodic-Other



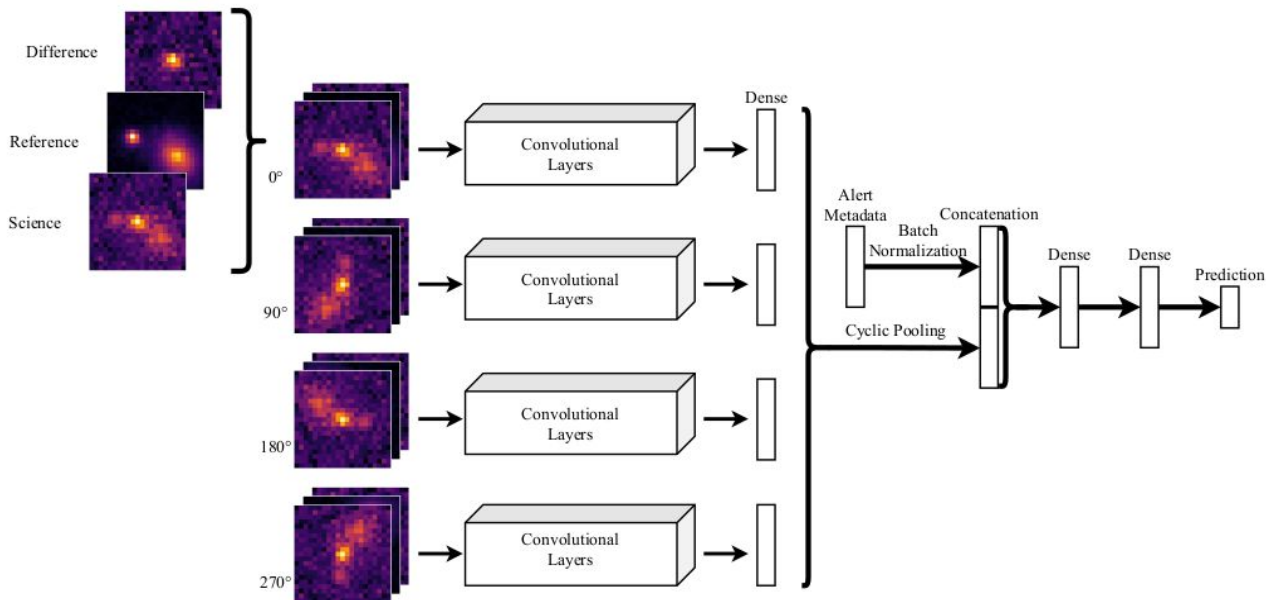


# Stamp Classifier vs Light Curve Classifier





# Stamp based classifier (Carrasco-Davis et al. 2020)

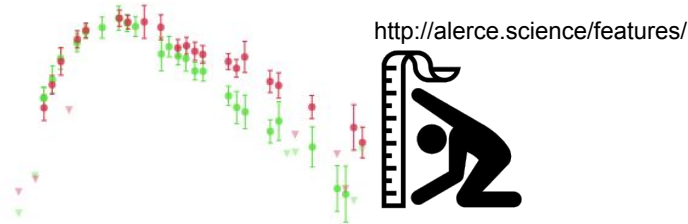


Layer	Layer Parameters	Output Size
Input	-	$21 \times 21 \times 3$
Zero padding	-	$27 \times 27 \times 3$
Rotation augmentation	-	$27 \times 27 \times 3$
Convolution	$4 \times 4, 32$	$24 \times 24 \times 32$
Convolution	$3 \times 3, 32$	$24 \times 24 \times 32$
Max-pooling	$2 \times 2, \text{stride } 2$	$12 \times 12 \times 32$
Convolution	$3 \times 3, 64$	$12 \times 12 \times 64$
Convolution	$3 \times 3, 64$	$12 \times 12 \times 64$
Convolution	$3 \times 3, 64$	$12 \times 12 \times 64$
Max-pooling	$2 \times 2, \text{stride } 2,$	$6 \times 6 \times 64$
Flatten	-	2304
Fully connected	$2304 \times 64$	64
Rotation concatenation	-	$4 \times 64$
Cyclic pooling	-	64
Concat with BN <sup>a</sup> features	-	$64 + 26$
Fully connected	$90 \times 64$	64
Fully connected	$64 \times 64$	64
Output softmax	$64 \times 5$	$5 (n^{\circ} \text{ classes})$



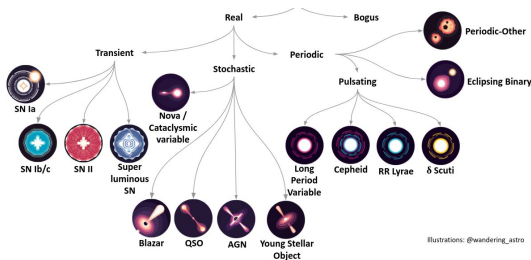
# Light Curve based classifier (Sánchez-Sáez et al. 2020)

Characterize



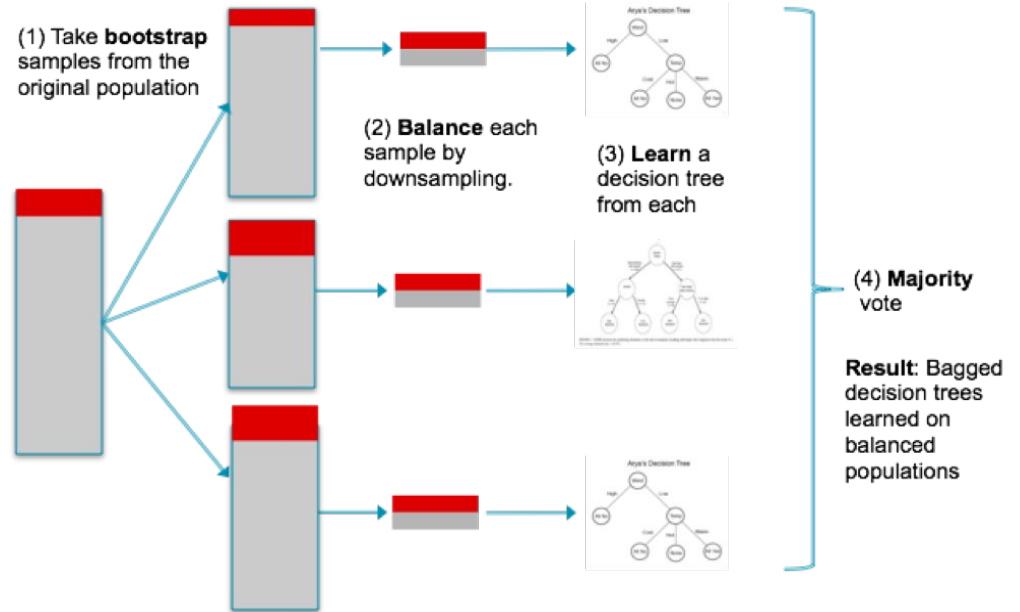
Nun+2015, Eyheramendy+2017/2018, Villar+2019

Define hierarchical structure



Sánchez-Sáez+2020

Train each classifier using bootstrap & balance.



Chen, Chao, Liaw, Breiman (2004)



# MIT Technology Review

## Hundreds of AI tools have been built to catch covid. None of them helped.

Some have been used in hospitals, despite not being properly tested. But the pandemic could help make medical AI better.

by **Will Douglas Heaven**

July 30, 2021

---

“Driggs highlights the problem of what he calls Frankenstein data sets, which are spliced together from multiple sources and can contain duplicates“.

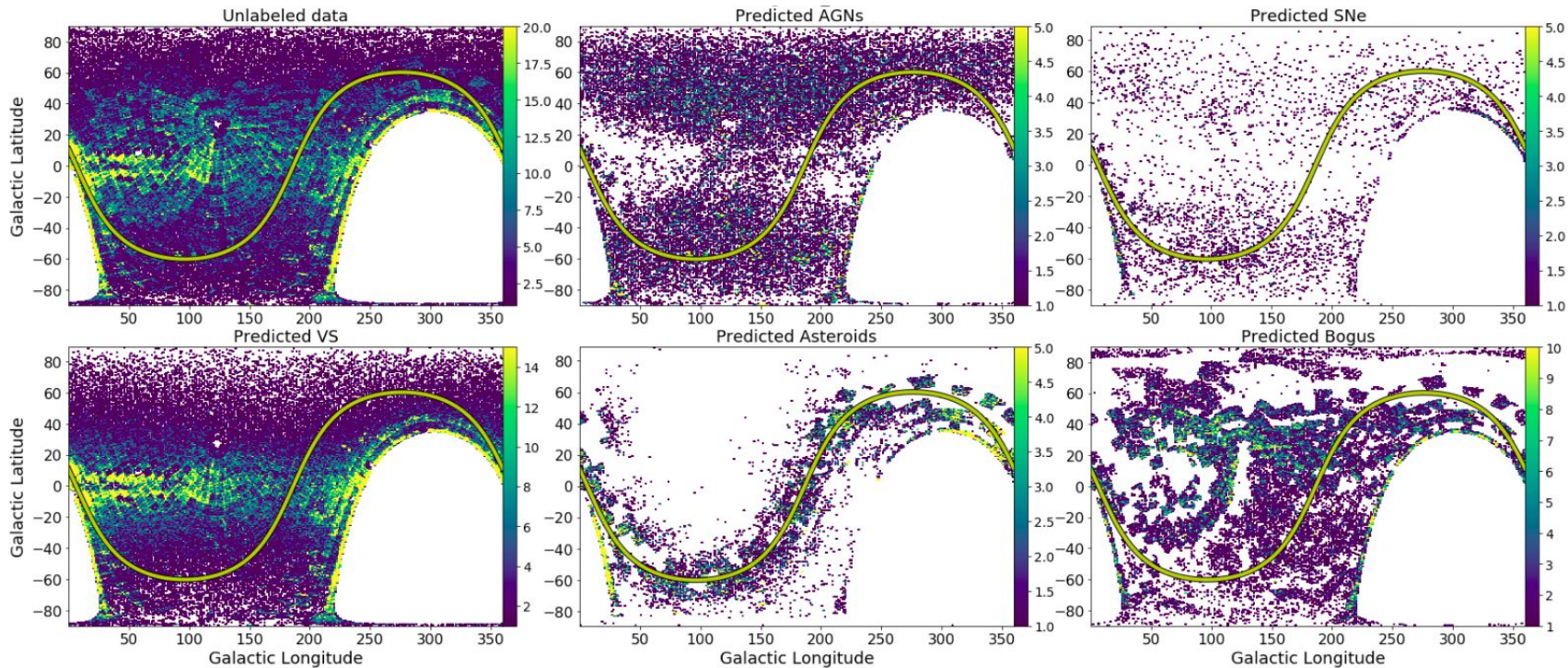
...many tools were developed either by AI researchers who lacked the medical expertise to spot flaws in the data or by medical researchers who lacked the mathematical skills to compensate for those flaws.

...Academic researchers have [few career incentives to share work or validate existing results](#). There’s no reward for pushing through the last mile that takes tech from “lab bench to bedside,” says Mateen.

ML in production >> ML in publication

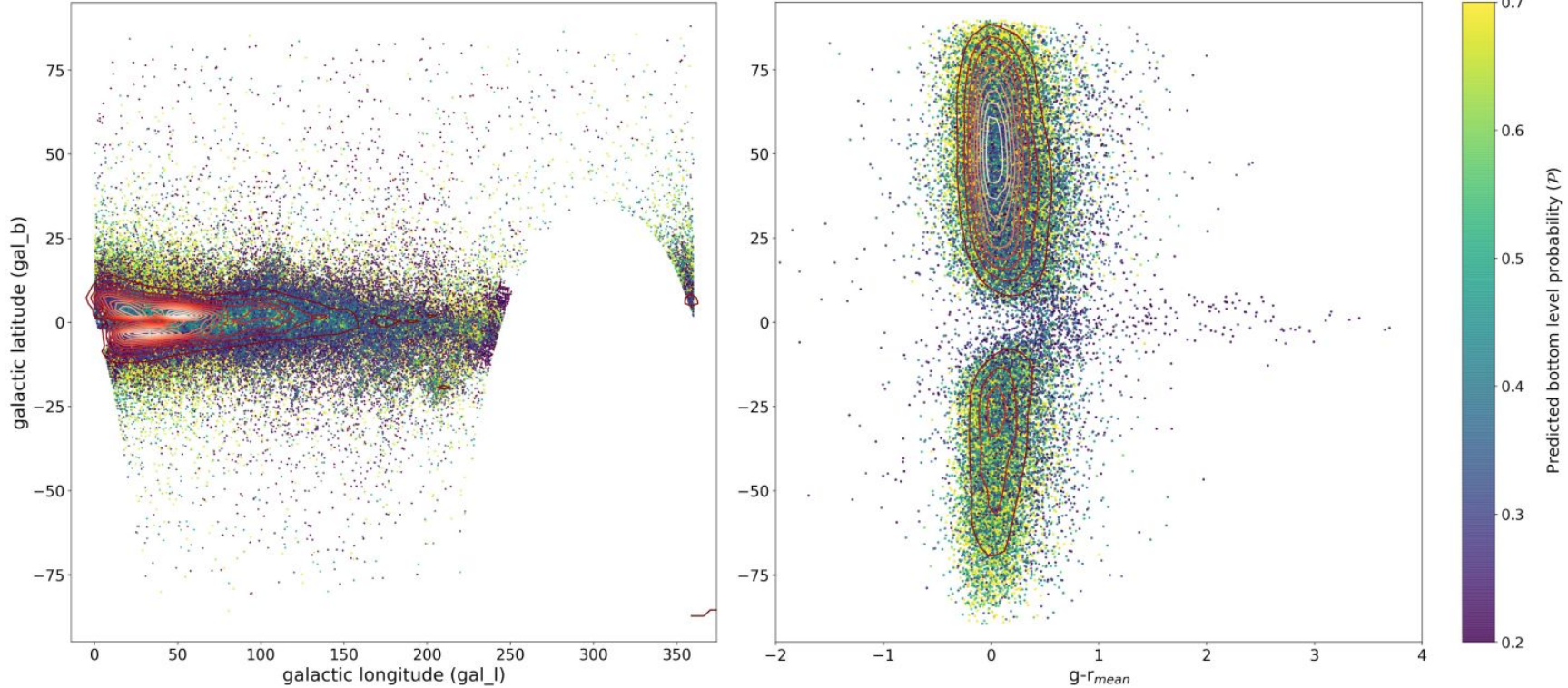


# Stamp Classifier: testing the unlabeled set



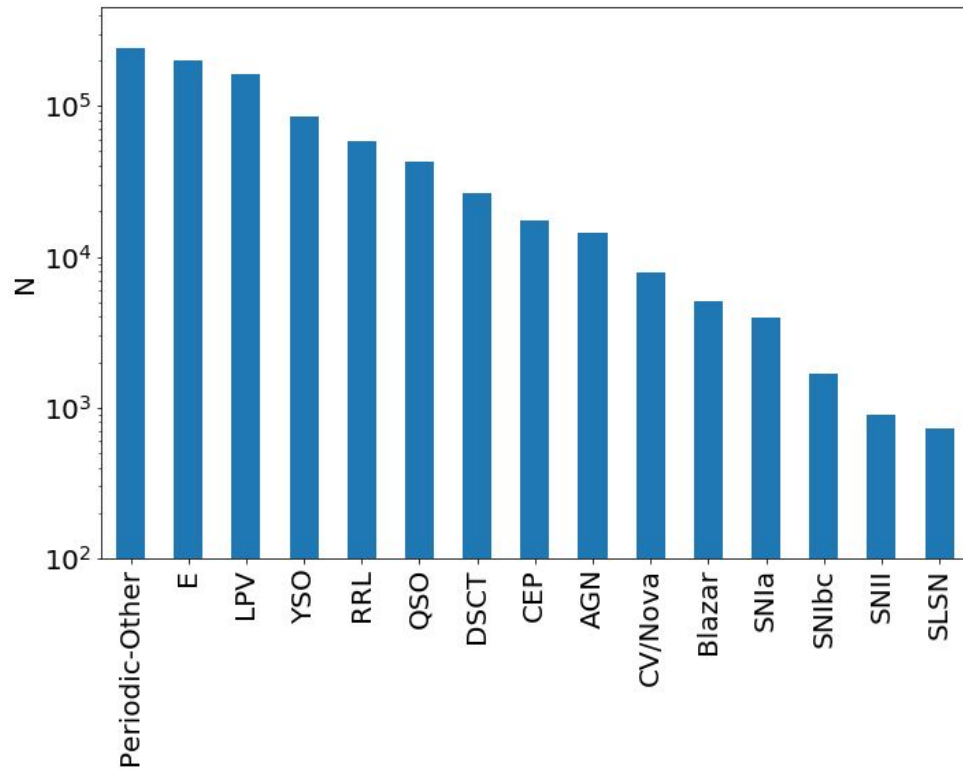
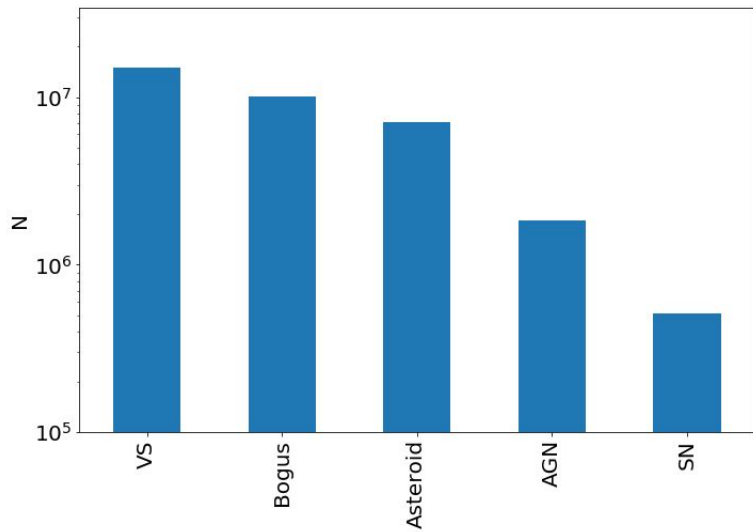


# Light Curve Classifier: testing the unlabeled set





## Testing the unlabeled set: class distribution







# ALeRCE Explorer



<https://alerce.online/>





# ALeRCE SN Hunter

The screenshot displays the ALeRCE SN Hunter web interface. At the top, the search results for 'ZTF21acpoujw' are shown, including its coordinates (RA: 171:01:55.0, DEC: 16:30:24.9) and a 'Discovery Range' gallery with three images. Below this is a large purple map of the discovery field. On the right, a table titled 'Top 100 SN Early Classified Candidates' is visible, listing various supernova candidates with their discovery dates, peak magnitudes, and redshifts.

Search	Discovery Date	Peak	z
ZTF21acpoujw	16111521-161055.6	16.0	0
ZTF21acpoujw	16111521-161055.6	16.0	0
ZTF21acpoujw	16111521-161055.6	16.0	0
ZTF21acpoujw	16111521-161055.6	16.0	0
ZTF21acpoujw	16111521-161055.6	16.0	0
ZTF21acpoujw	16111521-161055.6	16.0	0
ZTF21acpoujw	16111521-161055.6	16.0	0
ZTF21acpoujw	16111521-161055.6	16.0	0
ZTF21acpoujw	16111521-161055.6	16.0	0
ZTF21acpoujw	16111521-161055.6	16.0	0

<https://snhunter.alerce.online>



# ALeRCE Watchlist

<https://watchlist.alerce.online>

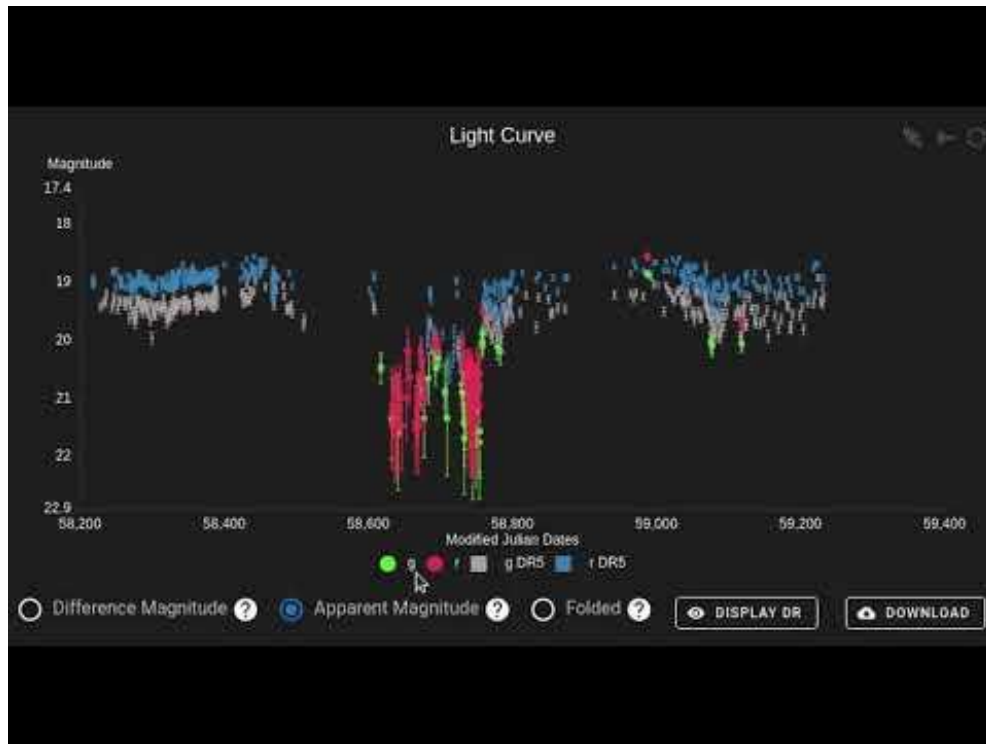
The screenshot shows the ALeRCE Watchlist web application interface. The top navigation bar includes the application name "ALeRCE Watchlist", a "HOW TO" link, and a "LOGOUT" link. The main content area is divided into two sections: "My Watchlists" on the left and "Targets" on the right. The "My Watchlists" section shows a single watchlist named "Watchlist: DICT1023". The "Targets" section is currently active and displays a table of targets. The table has columns for "Name", "Ip", "Port", "Status", "Matches", and "Action". Three targets are listed: ZTF21aciodbt, ZTF21acipnrf, and ZTF21acioda. Each target has a status of "1", a number of matches, and an action icon. A "NEW TARGET" button is located in the top right corner of the "Targets" section. The bottom of the "Targets" section shows a pagination control for "Showing page 11" out of "10 of 1".

Name	Ip	Port	Status	Matches	Action
ZTF21aciodbt	123.0727058	64.7748497	1	6	[Action Icon]
ZTF21acipnrf	168.2117754	88.2410852	1	1	[Action Icon]
ZTF21acioda	5.8887688	19.7984992	1	2	[Action Icon]



# ZTF data release integration & outlier detection

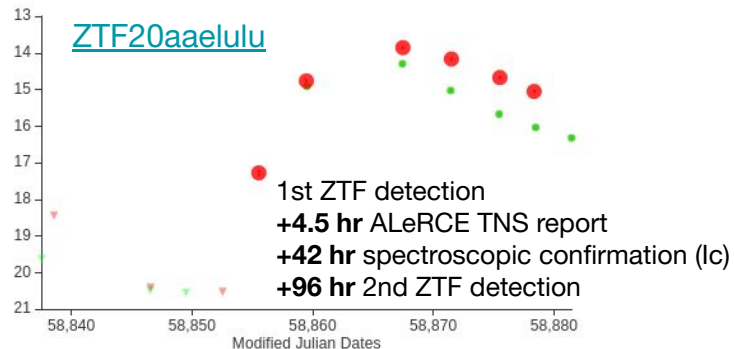
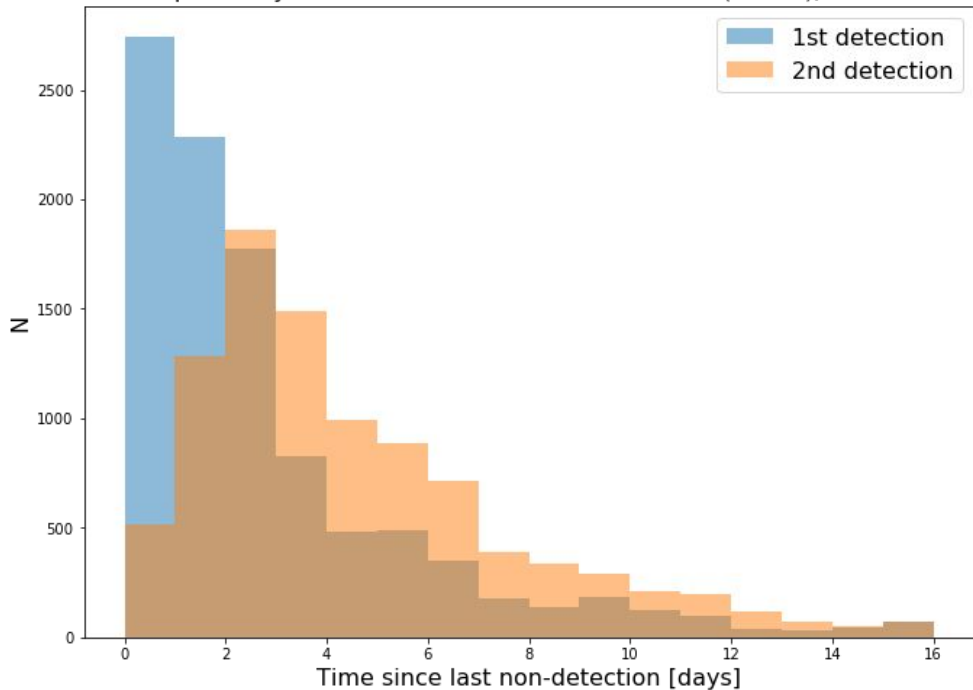
**Integration with ZTF Data Releases.**  
Enables **outlier** exploration: **changing state AGN as outliers**  
(Sánchez-Sáez+2021b) and study of **outlier detection algorithms**  
(Pérez-Carrasco, *in preparation*).



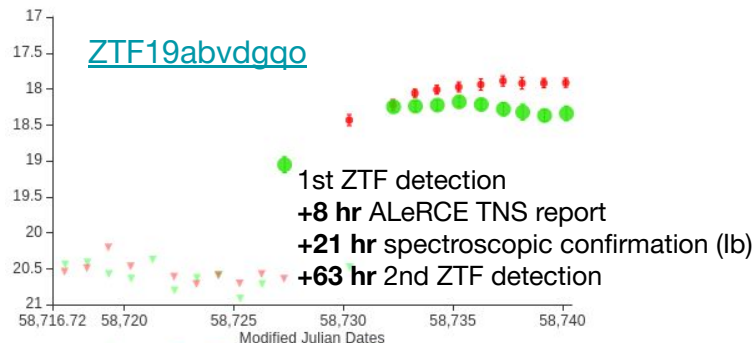


# Single stamp classifier = faster follow-up

SNe reported by ALerCE with at least two detections (10826), 2021/11/15



● g ● r ▼ g non-detections ▼ r non-detections



● g ● r ▼ g non-detections ▼ r non-detections



# Example: SN

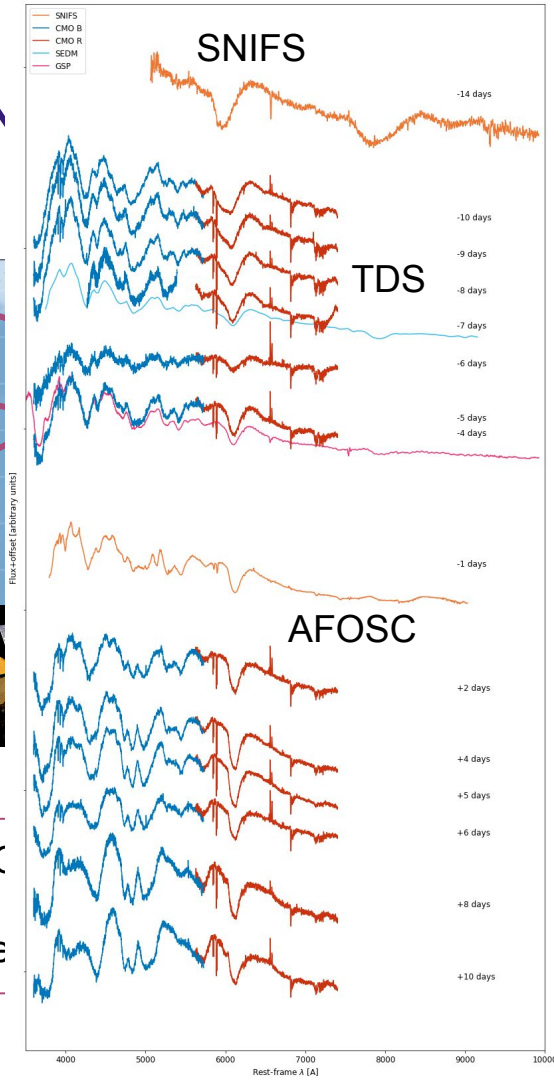
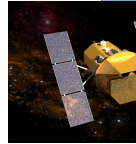
ZTF

ATLAS

SNIFS @ UH88:  
2 epochs (spec)

Swift/UVOT:  
15 epochs  
(phot)

Schmidt Tel & AFC  
@ Asiago:  
8 (phot) & 1 (spec) e



1 TDS  
in Observatory  
(spec) epochs


+ Global Supernova Project (GSP):  
11 (phot) & 1 (spec) epochs







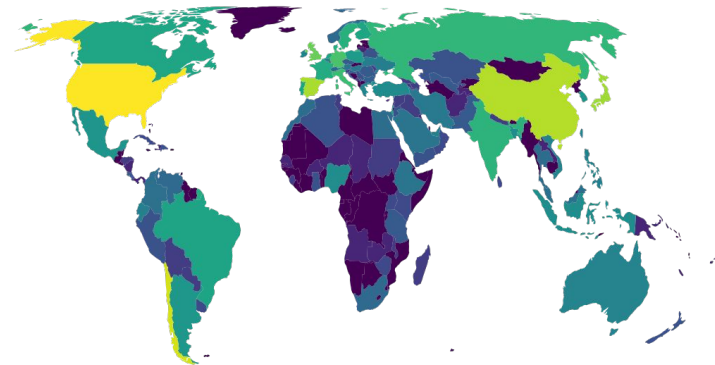
# ALeRCE user community

ALeRCE tutorial workshop and LSST Enabling Science Broker Workshops (with ).

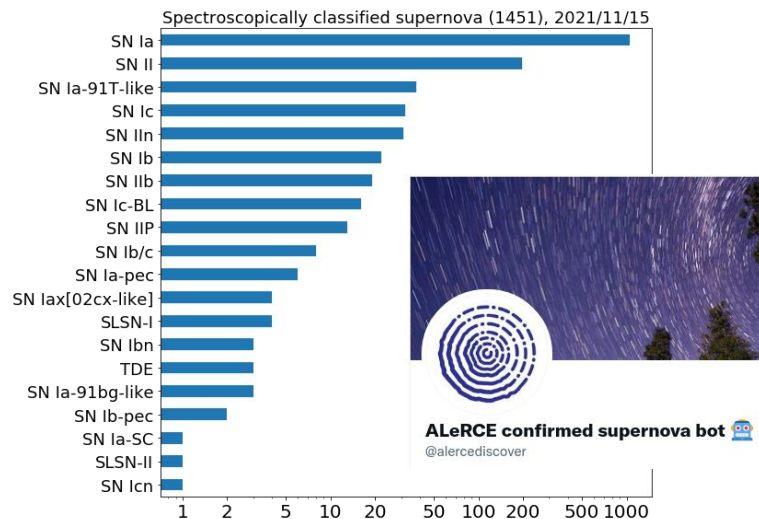
Feedback from **science collaborations**, **notebooks** focused on **science cases**, 76 recordings available!

<http://workshops.alerce.online/>

Services								
Crossmatch		✓	✓	✓	✓	✓	✓	
Follow up marshal		✓	✓	✓	✓	✓	✓	
Watchlists	✓	✓	✓		✓	✓	✓	
Classification	✓	✓		✓	✓	✓		
Features		✓	✓		✓	✓		
Outliers, anomalies			✓	✓		✓		✓
Data releases	✓				✓		✓	
Reproducibility				✓				✓
Multimessenger streams				✓				



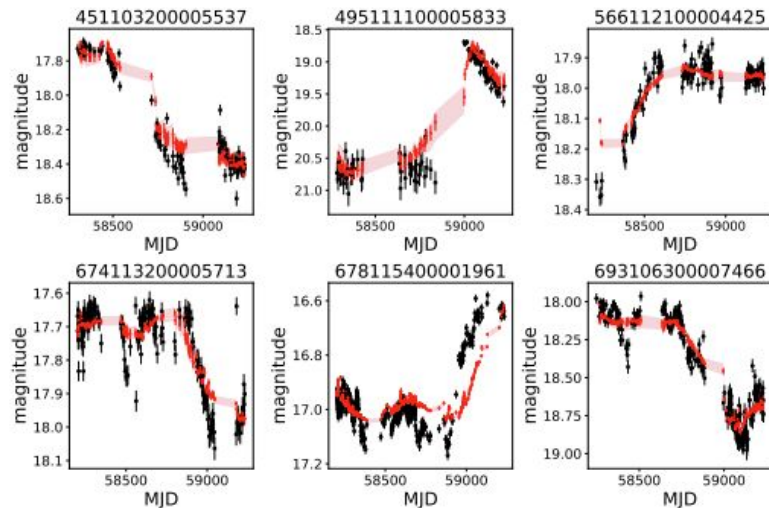
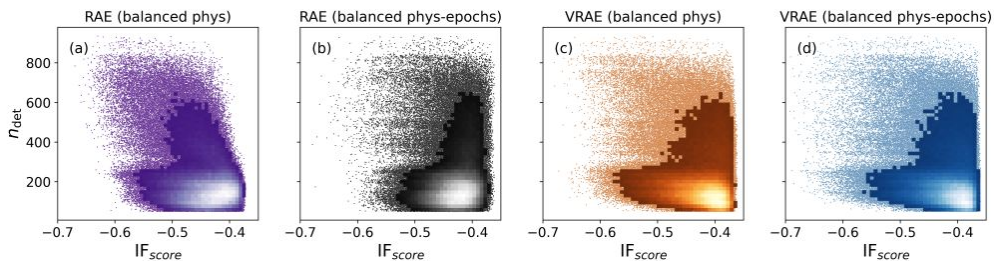
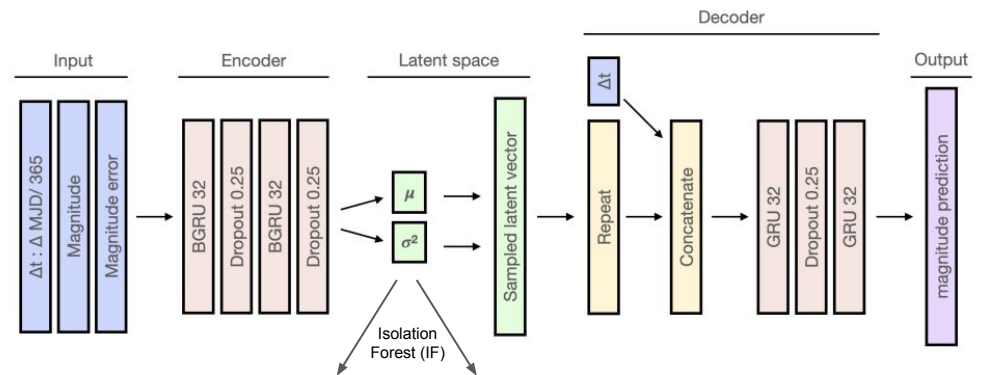
User community from 125 countries!



1451 SNe reported by ALeRCE spectroscopically confirmed

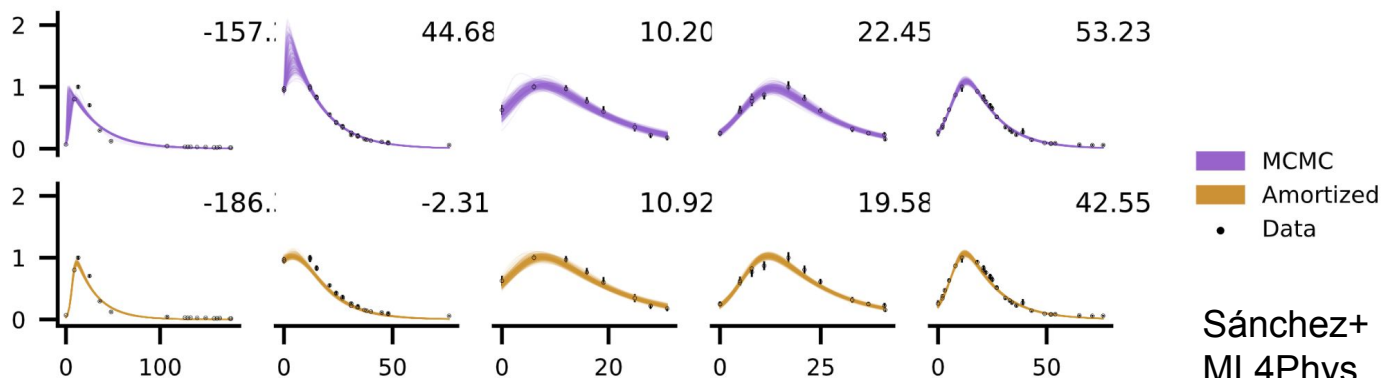
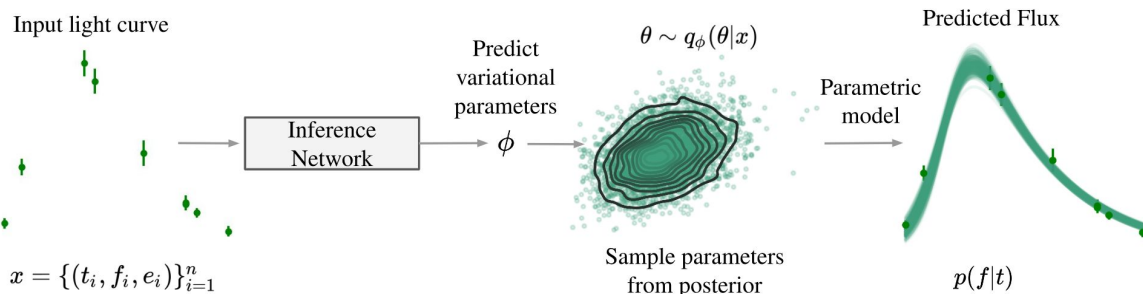


# Changing-state AGNs in massive datasets



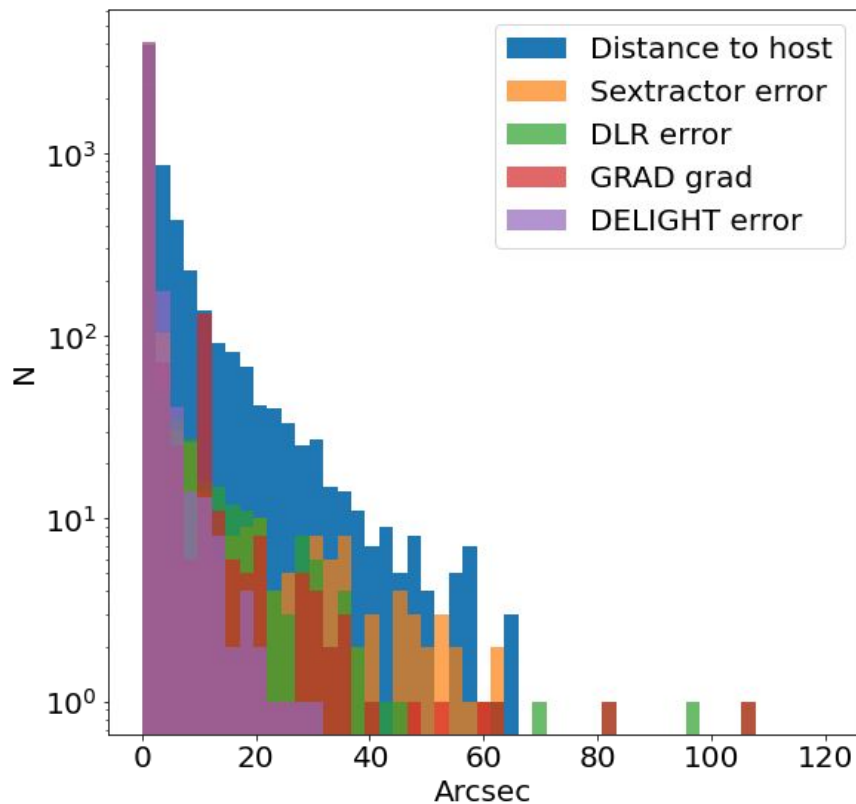
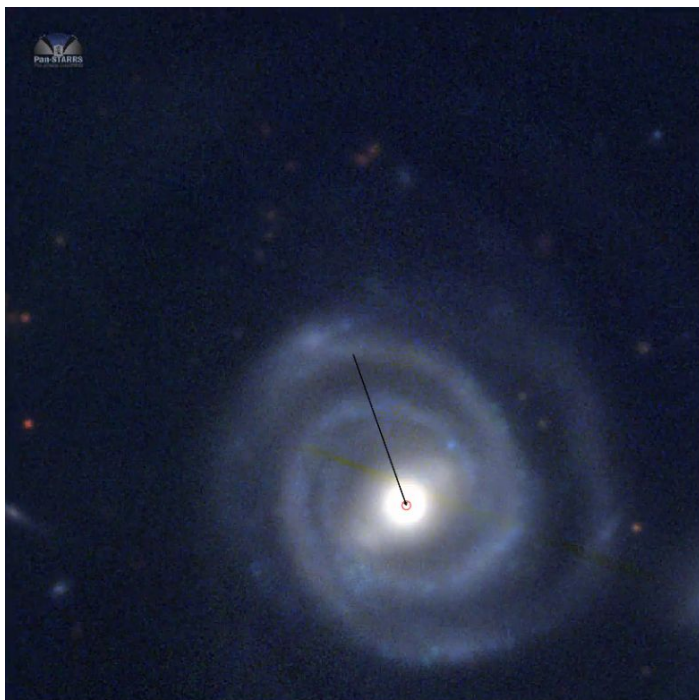


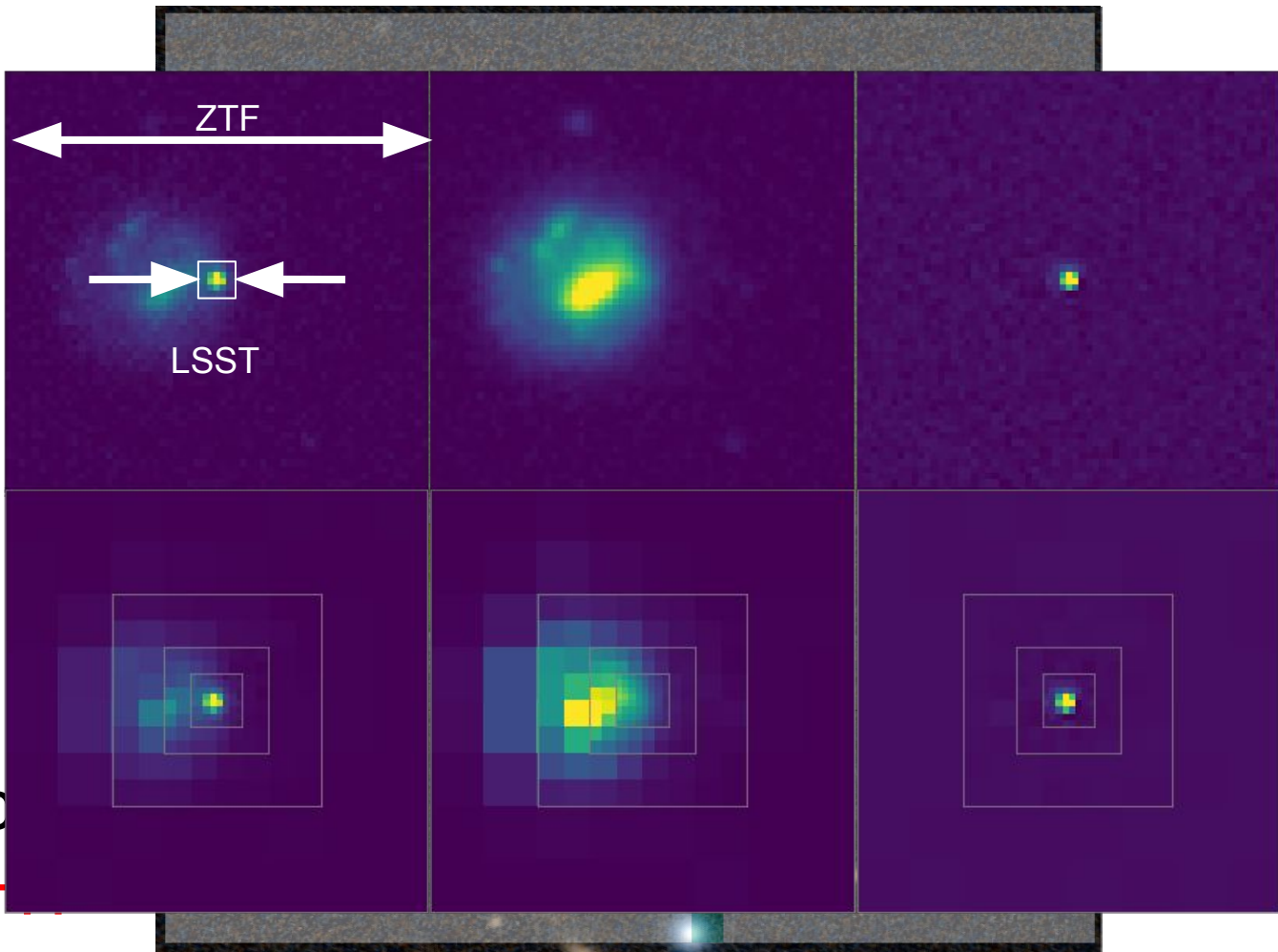
# Amortized variational inference for fast posteriors





# DELIGHT: Deep Learning Identification of Galaxy Hosts in Transients

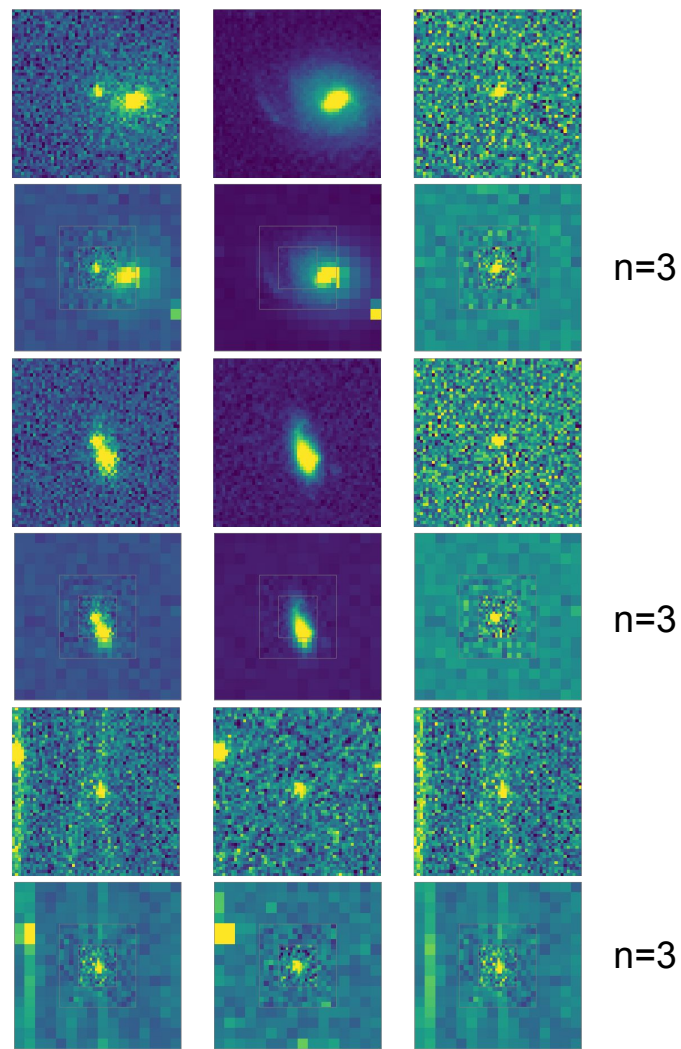
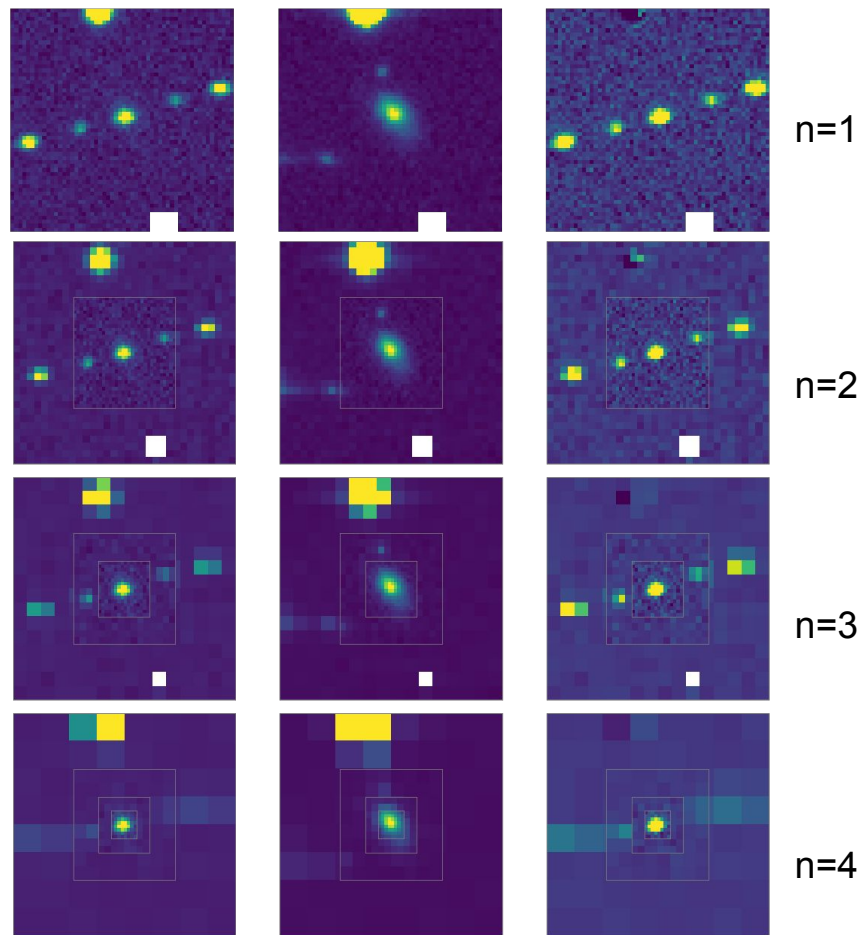




Sextant  
DELIGHT



# Examples





# Under construction: multi-stream broker

## Bridging ZTF - ATLAS

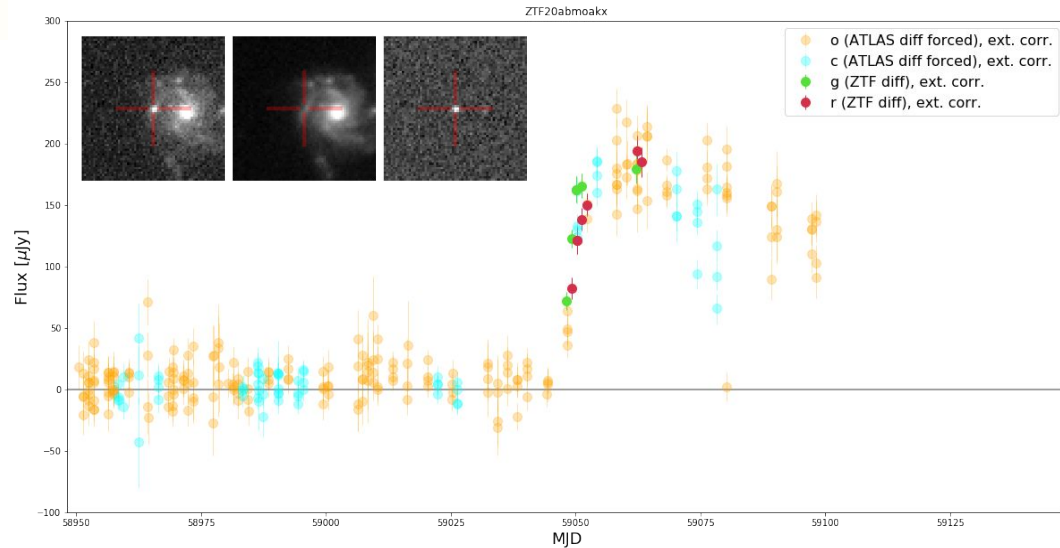
Collaboration with ATLAS and Lasair,  
building **multi-stream broker**



## TOM connected

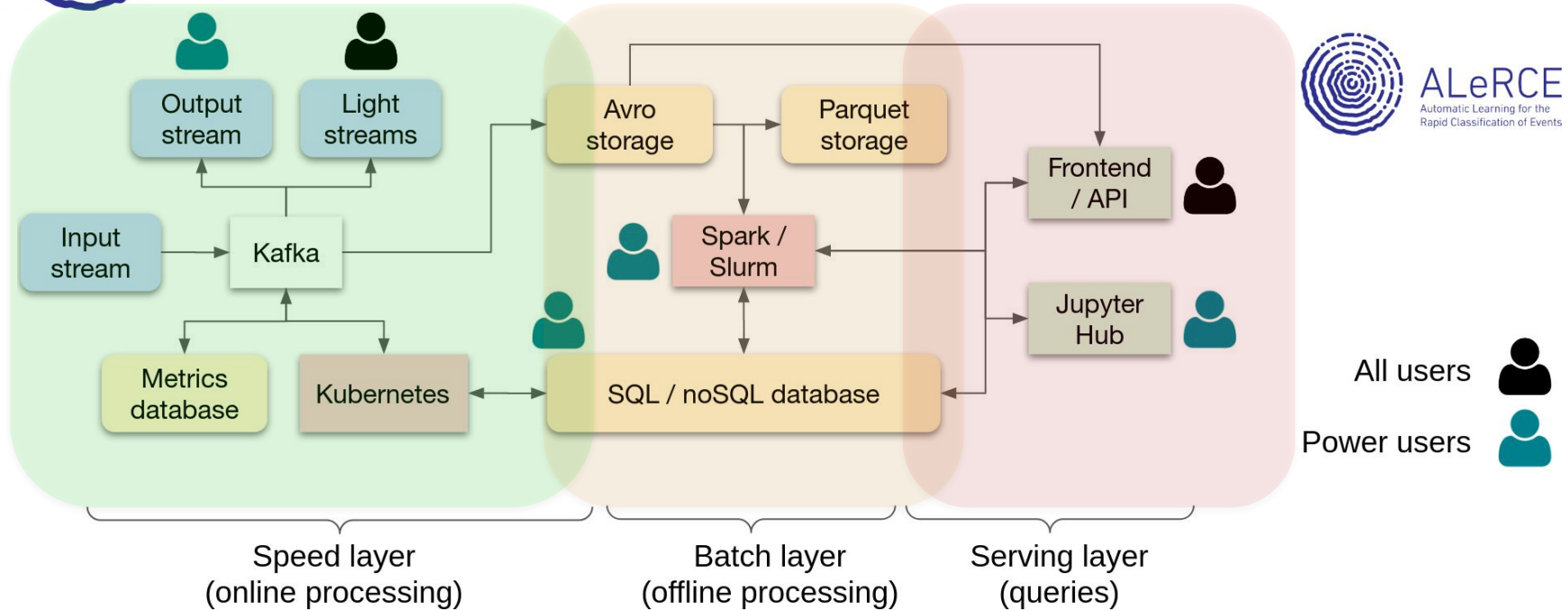
Custom frontend at  
<https://tom.alerce.online/> .

>600 observations submitted.





# ALeRCE infrastructure



ALeRCE  
Automatic Learning for the  
Rapid Classification of Events

All users

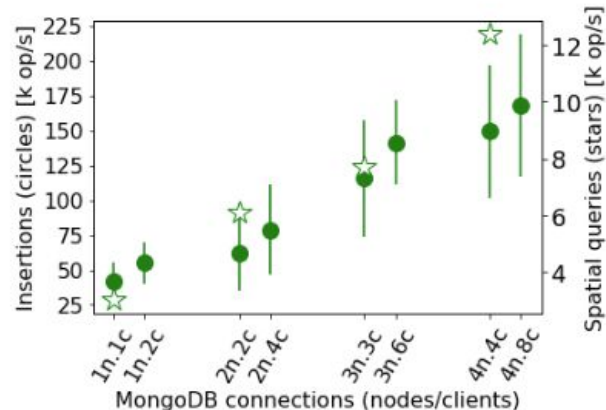
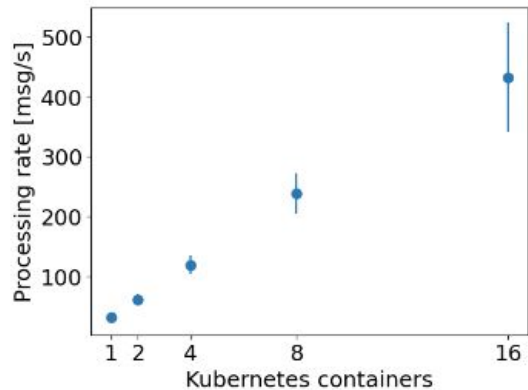
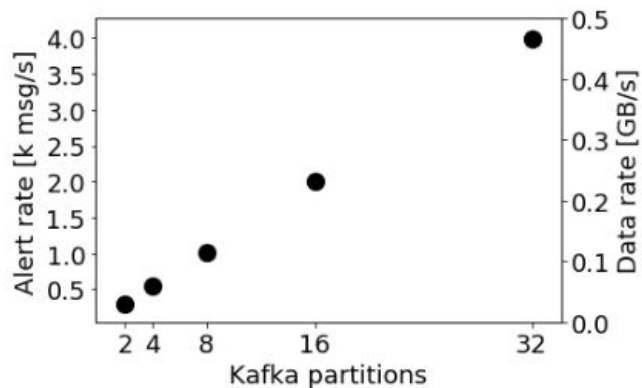


Power users





# The future: Vera C. Rubin



Simulating alert ingestion with Kafka, alert processing using Kubernetes, and spatial queries using MongoDB at LSST level.

## Explorer

Search and visualize  
<https://alerce.online/>

## Notebooks

SNe, Variable stars, AGN

## ALeRCE Watchlist

Targets	Matches			
Name	Ra	Dec	Object	Date
Mrk590	2.2426667	-0.7666667		
NGC1068	2.7113056	-0.0133333		
Mkn609	3.4236667	6.1438889		
NGC1365	3.5601111	-36.1402778		
1H0419-57	4.4338278	-57.2002778		

## LC Classifier

15 class taxonomy  
 1.4 million classifications

## SN Hunter

Find recent SNe  
<https://snhunter.alerce.online/>

## Pipeline

Modular design

## VERA C. RUBIN

Speed+Batch+Serve layers

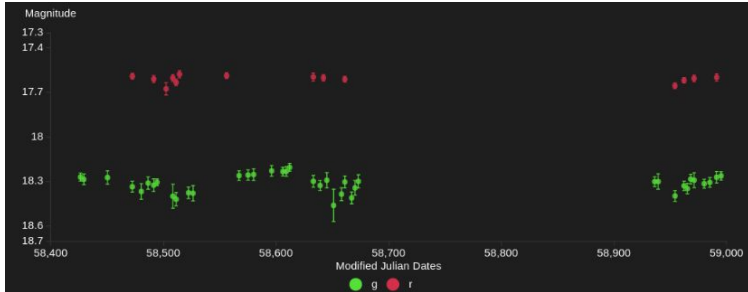
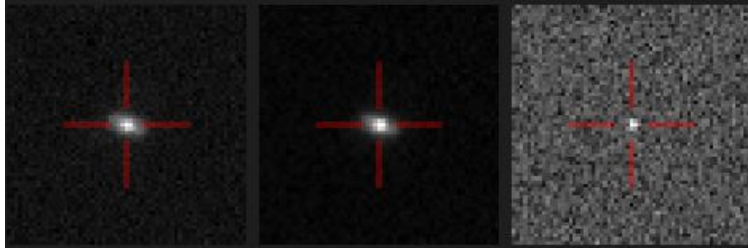


<http://alerce.science/>





# Light curve correction



Difference magnitude is given. If source in reference image, its magnitude is also provided.

We reconstruct total flux when possible, using formulae below:

$$m_{\text{corr}} = -2.5 \log_{10} \left( 10^{-0.4 m_{\text{ref}}} + \text{sgn} 10^{-0.4 m_{\text{diff}}} \right)$$

$$\delta m_{\text{corr}} = \frac{\left( 10^{-0.8 m_{\text{diff}}} \delta m_{\text{diff}}^2 \left[ -10^{-0.8 m_{\text{ref}}} \delta m_{\text{ref}}^2 \right] \right)^{0.5}}{10^{-0.4 m_{\text{ref}}} + \text{sgn} 10^{-0.4 m_{\text{diff}}}}$$

We flag points which show inconsistencies, e.g. reference source not reported after change of reference.

Magnitude statistics and light curve features also computed.