SYNERGIES IN THE EXPLORATION OF THE EXTREME UNIVERSE

for researchers

“Alert Brokers for Astrophysical Surveys: What is in it for CTA and how to engage”

Gautham Narayan
Assistant Professor at the University of Illinois at Urbana-Champaign
Deputy Analysis Coordinator of the Rubin Observatory’s Dark Energy Science Collaboration
1. How Astrophysical Surveys Are Evolving
DEEP LENS SURVEY
37 billion stars and galaxies (10 year survey)

10 million alerts, 1000 pairs of exposures/night
THIS IS THE CORE OF AN ALERT

THIS IS WHAT THE DATA LOOK LIKE WHEN YOU SUBTRACT AWAY MOST OF IT
LSST watches the changing sky, discovering the deaths of stars
LSST’S ALERT RATE OUTSTRIPS ALL OUR FOLLOWUP RESOURCES - COMBINED

8m Class Telescope

Primary Mirror Diameter

Field of View

Z. Ivezic,
The Bulk of LSST alerts users will be LSST DESC, TVSSC and AGNSC (in that order)

**Transients**

- Rapid photometric & spectroscopic follow-up: nature of the progenitors (outermost layers) & explosion physics (ejecta structure).
- Short-lived transients (GWs, GRBs)
- Cosmological distance rulers.
- Rare populations of events?

**Variable stars**

- Rapid photometric & spectroscopic follow-up: low mass *microlensing* events, *changing mode* stellar pulsators, rapid reaction to *eclipsing events*, *eruptive* events
- Analysis of large populations of events: study Milky Way structure & formation.

**Active Galactic Nuclei**

- Rapid photometric & spectroscopic follow-up: changing look AGNs, reverberation mapping studies
- New populations of faster, redder, dimmer AGNs
- Detection of intermediate mass black holes, tidal disruption events
Kilonova

The galaxy NGC 4993

Size of full moon for comparison

The Fermi satellite detects a gamma-ray burst from this area of the sky

The LIGO and Virgo detectors triangulate a gravitational wave signal from this area of the sky

3 KILONOVA

Las Cumbres Observatory

The galaxy NGC 4993
RARE AND INTERESTING TRANSIENTS IN THE ALERT STREAM: NEEDLE IN A NEEDLESTACK
TRADITIONAL ARCHIVES

Access to raw/de-trended data products after some proprietary period. Limited query, no compute.
SDSS - CATALOG QUERIES VIA SQL

Community access to SDSS images, spectra, catalogs in data releases + nice interfaces and API. Query but not compute.
WE NEED TO FIND A BETTER WAY TO FILTER

THE REAL ISSUE IS VOLUME AND RATE

WE CAN'T AFFORD TO MISS THE RAREST OF THE RARE
The key challenge for the next generation of surveys is not just taking vast quantities of data, but how to enable the community to use it.

For the time-domain, this implies real-time data streams from heterogeneous sources.
2. WHAT ALERT BROKERS CAN DO
WHAT DO WE NEED?

- Something to sift through heterogenous alert streams in real-time
- Characterize and classify events
- Identify outliers
- Prioritizes events for follow-up
- Actively learns from the follow-up
- All while providing a search, filtering and compute service to the community
Alert Distribution and Brokers

Alerts are delivered to the community brokers that filter and classify events to enable realtime science.

Community Brokers

- Software developed independently of LSST to receive, characterize and/or redistribute Alerts.
- Added functionality such as, filtering, photometric classification & cross-matching with other surveys
- Limited number selected by a proposal process receive the full realtime stream

LSST Alert Filtering Service

- Limited capacity service provisioned through the LSP; simple filters on alert packet contents only
Alerts for ZTF now and soon for LSST use an Apache Kafka queue.

Alerts are formatted using Apache Avro (effectively serialized JSON, with a schema).


Very much a successor to VOEvent.

SCiMMA group is adding an IAM layer to Kafka with HopSkotch: https://scimma.org/projects.html

HOW CTA CAN ENGAGE 1: START TO WORK WITH SCiMMA/LSST/ZTF/LVK/SNEWS ON ALERTS
ANTARES manages alert streams, adding contextual information, characterizing events, ranking & distributing.

Write your filters for complex, targeted processing, or broad analysis of large datasets - do YOUR science!

Correlate optical, GW & high-energy particle/neutrino triggers - multi-messenger astrophysics.

Active with ZTF! Sign up! [https://antares.noirlab.edu/](https://antares.noirlab.edu/)
The next few slides cover the same material in the live demo in static form for those looking at the slides. In each case, I’ve highlighted some of ANTARES’ capabilities followed by some science cases that can use them.
You can think of ANTARES as a search engine for astrophysical variability. You can coneselect, cross-match against existing catalogs at [https://antares.noirlab.edu/catalogs](https://antares.noirlab.edu/catalogs)

Even require objects that have “tags” attached to them at [https://antares.noirlab.edu/tags](https://antares.noirlab.edu/tags)

Or create a private watch list to monitor activity for your own sources at [https://antares.noirlab.edu/watch-lists](https://antares.noirlab.edu/watch-lists)
Nuclear Activity: AGN and TDEs

- Identify by location (currently using Van Velzen criteria)
- Multiwavelength catalogs can help distinguish AGN from TDE
- Multiwavelength variability would be useful

RAPID classifier identifying ZTF18abxftqm as TDE, Muthakrishna et al. 2019
Science Use Cases

Watchlists

- Users upload csv file
- Direct notification when your object of interest alerts
- Slack configuration controls intrusiveness of notifications
- Watch all known strong lensing systems for transients, e.g.
Level 2 Filters

After alerts are ingested, aggregated, and associated with catalogs, ANTARES runs the L2 filters. The purpose of the L2 filters is to detect the interesting science alerts based on custom criteria.

Tags are set by “filters”
https://antares.noirlab.edu/tags

ANTARES already has several filters, both created by our staff, and by community users (you!)

YES! We run YOUR code as part of OUR pipeline!

You can develop your own filters using the DevKit Jupyter notebook on AstroDataLab:
https://datalab.noirlab.edu/

and there’s even documentation:
https://noao.gitlab.io/antares/filter-documentation/devkit/index.html#devkit

Tags are set by “filters”
https://antares.noirlab.edu/tags

ANTARES already has several filters, both created by our staff, and by community users (you!)

YES! We run YOUR code as part of OUR pipeline!

You can develop your own filters using the DevKit Jupyter notebook on AstroDataLab:
https://datalab.noirlab.edu/

and there’s even documentation:
https://noao.gitlab.io/antares/filter-documentation/devkit/index.html#devkit
Science Use Cases

Longer-timescale Transients/Variables

- Many classes don’t require immediate response
- Searchable annotated archive provides resource to discover these
- Useful for developing filters for shorter-timescale objects
New Solar System objects are on-project task
Known Solar System objects flagged in the alert stream can be redirected to a moving-object broker
Already doing this with ZTF stream and SNAPS team
Filter on streaked sources
Science Use Cases

Enabling Multi-Messenger Astronomy

- Automatic retrieval of LIGO/Virgo skymaps
- Associate all subsequent alerts within recent skymaps
- Filter using distance and other features

LIGO/Virgo Gracedb
import antares_client

from astropy.coordinates import Angle, SkyCoord

center = SkyCoord("227.285d 67.222d")
radius = Angle("0.25d")

locusid=[]
locus_gr=[]
locus_r=[]

print("#Locus, RA, Dec, Num_alerts")
for locus in antares_client.search.cone_search(center, radius):
    if locus.properties['num_mag_values'] > 30:
        print("https://antares.noirlab.edu/loci/lookup/%s" % (locus.properties['ztf_object_id']), locus.ra, locus.dec, locus.properties['num_mag_values'], locus.alerts[-1].properties['ztf_srmag1'] - locus.alerts[-1].properties['ztf_srmag1'])
        locusid.append(locus.properties['ztf_object_id'])
        locus_gr.append(locus.alerts[-1].properties['ztf_srmag1'] - locus.alerts[-1].properties['ztf_srmag1'])
        locus_r.append(locus.alerts[-1].properties['ztf_srmag1'])

#Locus, RA, Dec, Num_alerts
https://antares.noirlab.edu/loci/lookup/ZTF18aapnxp 227.4650415090125 67.23184959776788 45 -0.10289955139160156 20.2005048828125
https://antares.noirlab.edu/loci/lookup/ZTF18avahko 227.26304767490103 67.1691137945075 37 0.028499603271484375 19.3388961246758
https://antares.noirlab.edu/loci/lookup/ZTF18aapmbqs 227.3569277820102 67.203745733418 74 0.277599347167969 18.9530064086914
https://antares.noirlab.edu/loci/lookup/ZTF18aqwgh 227.352061211087 67.1009379691194 98 0.6824989318847656 19.34210014332617
https://antares.noirlab.edu/loci/lookup/ZTF18aapscwd 227.6100967449605 67.124552552642 41 0.7966093471679675 19.395099639825478
https://antares.noirlab.edu/loci/lookup/ZTF18aqmvlf 226.96067721846745 67.15751626751334 44 -0.5040005432128906 19.96330704956055
https://antares.noirlab.edu/loci/lookup/ZTF18aaoexp 227.3073308554661 67.25926222542068 376 1.07369951171875 15.774299621582031
https://antares.noirlab.edu/loci/lookup/ZTF18aapmlzb 227.58436607491905 66.99700828060137 181 0.47480010986328125 18.33309936523475
https://antares.noirlab.edu/loci/lookup/ZTF18aapklkb 227.105624585155 67.1582211012635 88 0.1955013275146484 19.472299575805664
https://antares.noirlab.edu/loci/lookup/ZTF18aatfbe 227.64738291707317 67.0454438 41 0.4519996643066406 14.103400230407715
https://antares.noirlab.edu/loci/lookup/ZTF20aabqwer 227.70058449230774 67.21825292417583 91 0.4307003021240234 18.2782993166504
As an example, instead of looking at sources singly, you can write an SQL query as you might with SDSS CASJobs.

Submit it to ANTARES, and get all matching objects from the real-time alert stream (NO WAITING FOR A DATA RELEASE!)

You can make plots like a CMD interactively to select outlying sources.

This notebook is available here: https://github.com/broker-workshop/tutorials/blob/main/ANTARES/ExploringVariabilityWithANTARES.ipynb
REAL-TIME FOLLOWUP WITH TOMS

Convince a TAC to give you time

Find object of interest in the stream

Create a followup observation request and submit

This notebook is available here:
We can also see the observation request at the LCO observation portal:

submitted Observation Requests

Programmatic Observation

lee

TOM2020A-012

PENDING

2020-10-09 16:4...

1 1 0 0

Telescope availability history?

<table>
<thead>
<tr>
<th>Telescope</th>
<th>-3 days</th>
<th>-2 days</th>
<th>-1 day</th>
<th>Today</th>
</tr>
</thead>
<tbody>
<tr>
<td>Siding Spring 0.4m B</td>
<td>64</td>
<td>6</td>
<td>80</td>
<td>98</td>
</tr>
<tr>
<td>Siding Spring 2m</td>
<td>59</td>
<td>6</td>
<td>95</td>
<td>98</td>
</tr>
</tbody>
</table>
Transients on Demand

- Common transients don’t all need follow up
- Targeted/scheduled programs can get transients as needed
- Filters can be flexibly scheduled to accommodate a variety of needs

Pan-STARRS Type Ia Supernovae, Jones et al. 2018
The Python Client + TOMs integration means you can roll a full real-time followup pipeline easily

- Find an interesting source in CTA data
- Send it out in your alert stream
- Use the client to see if it is associated with anything in optical or gets tagged if it passes your filter criterion
- Report to TNS automatically
- Use TOM Toolkit to schedule followup

https://www.wis-tns.org/object/2021pkt
HOW CTA CAN ENGAGE 2-4:  

IF YOU HAVE A CATALOG OF SOURCES ALREADY, YOU CAN START MONITORING THEM IN OTHER BANDS ALREADY  

PATHFINDER FACILITIES ARE GREAT SOURCES OF ALERTS FOR TESTING! REACH OUT TO ANTARES IF YOU WANT TO PUBLISH THEM!  

SCIENCE USE CASES IMPLICITLY DEFINE FILTERS – WE CAN HELP DEVELOP THEM!
3. USING BROKERS TO OPTIMIZE SURVEYS
ML FOR CLASSIFYING SOURCES

If you are looking at the slides PDF, this video can be found at: https://www.youtube.com/watch?v=jgO0JU_I5-s
PLAsTiCC Version 1  (Dec. 2018 - Feb 2019)

- Public $25k Kaggle challenge for photometric classification of time-domain sky (15 models, 1 million new SEDs, unrepresentative training sample)
- Data: 3M VRO-simulated $ugrizY$ lightcurves
- Primary goal: setup massive time-domain simulation infrastructure, **jump start**
**ML photometric classification efforts**

(Villar+2021)

(Hložek+2020)
If you believe the outcome of PLAsTiCC, then MMA source discovery will not be a problem!

<Narrator>: But it will be a problem!

LSST Cadence has a median intranight gap of ~4 days across all filters, and ~10 days in any single filter.

Classifiers only as good as their training data and their loss function.
(Some known) shortcomings of v1

What tells you that 21jap at right is real?

Simply, flat light curves with full phase coverage aren’t a realistic model for how the science collaborations will interact with LSST data.

*No host galaxy information, postage stamps, or alerts*

Surveys cannot afford to act independently
Goal for PLAsTiCC v2 (~September 2021)

To evaluate* real-time Broker performance on a realistic LSST alert stream.

**Broker Roles**
- Storing, processing, classifying alerts, informing follow-up
- Potential additional roles:
  - Collecting active source features
  - Maintaining source databases

**Alert Stream**
- Set of simulated LSST-like alert packets
- must preserve environmental correlations
- Should contain a representative sample of expected events

Iterative process to ensure compatibility between brokers and alert stream.

*metrics in progress!
RAPID: REAL-TIME AUTOMATED PHOTOMETRIC IDENTIFICATION = DEEP LEARNING TO IDENTIFY TRANSIENTS EARLY!

![Diagram](diagram.png)

**Dr.** Daniel Muthukrishna (Cambridge) uses the same tech in predictive text to predict light curve behavior.
Galaxies HOSTing Supernovae and other Transients

Alex Gagliano (UIUC, NSF Fellow) figures out correlations between transients and their hosts http://ghost.rubin.science/

- 16,228 SNe-host galaxy pairs: 78% of unique events reported on TNS/OSC.
- PS1,NED photometric & derived properties (color, redshift, radial moments)
v2: Simulating Transients & Hosts

**Pre-processing**
- GHOST
  - 16.5k SNe, host galaxies
  - Gagliano+2021
- cosmoDC2
  - DESC Synthetic Sky Catalog
  - arXiv:1907.06530
  - Korytov+2019

**Simulation**
- SNANA
  - Simulated transient photometry (with host galaxy properties)
  - arXiv:0908.4280
  - Kessler+2009
- EmpiriciSN
  - Simulated transient parameters
  - arXiv:1611.00363
  - Holoien+2016

**Validation**
- LIGO
  - Localized Event Skymaps
  - D. Chatterjee

**Brokers**
- ANTARES
- Pitt - Google
- ALeRCE
- LASAIR
- Fink
- AMPEL
- MARS
- INAF broker
- Fritz
- Babamul
- South African broker team
- NYU Anomalies
- SNAPS
- UW Genesis

**v2 Alert Stream**
El-Cid

Electromagnetic Counterpart Identification

Deep Chatterjee (UIUC, Illinois Survey Science Fellow, LVK EM Followup Team)
Performance on GW170817

DECam

MJD

Kn class

Other class
HOW CTA CAN ENGAGE 5:

EVERY SURVEY CREATES LARGE SIMULATIONS OF WHAT THEY EXPECT TO SEE TO TEST SURVEY DESIGN

THIS NEED NOT BE JUST INFRASTRUCTURE WORK, AND CAN REALLY ENGAGE THE COMMUNITY SCIENTIFICALLY – E.G. PLASTICCC

PLASTICCC V2 ALREADY INCLUDES LVK AND LSST. V3 COULD HAVE THE CTA SKY WE’D LOVE TO WORK WITH YOU!
RESEARCH PLATFORMS IN THE CLOUD ARE COMING
LSST and CTA both build on current-generation surveys to discovery many new sources.

Maximizing scientific utility from these experiments requires providing data in real-time.

Get away from just the concept of an archive with periodic data releases!

Alert brokers are a key tool for this and give the community ways to search, characterize, filter, and follow-up events.

They enable machine learning in real-time for astrophysical surveys.

There are already de facto standards for alert format and systems to stream them, as well as existing brokers - you don’t have to start from zero!!!

As CTA builds out it’s pipeline, and creates simulations for survey design, interact with broker teams and the scientific community and define interesting questions that also spur infrastructure development.

The best science will come from combining real-time streams from several experiments - challenge but also a tremendous opportunity.