

Taming the Fire Hose: Using ANTARES to Make Sense of the LSST Alert Stream

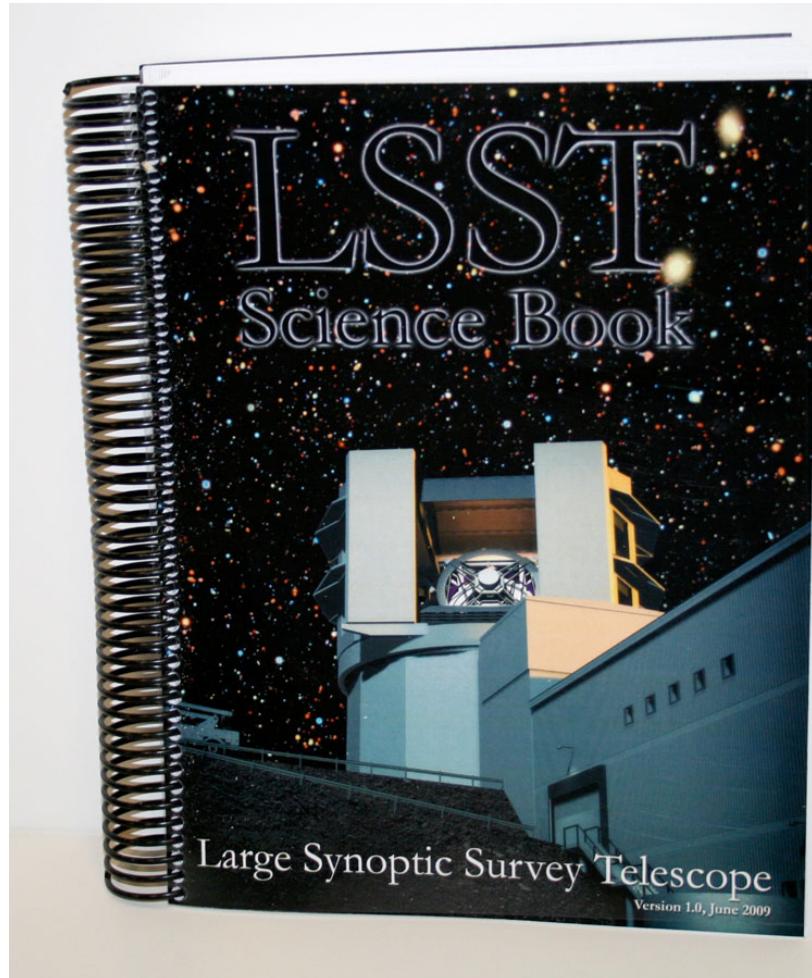
Tom Matheson



UA CS: Rick Snodgrass, John Kececioglu, Carlos Scheidegger, Zhe Wang, Shuo Yang, Jackson Toeniskotter, Clark Taylor, Eric Welch
NOAO: Abi Saha, Gautham Narayan
Catalina Sky Survey: Rob Seaman
LSST: Tim Jenness
NOAO REU: Tayeb Zaidi (Macalester)

LSST: Science Goals

- Milky Way structure (galactic archaeology, populations, streams)
- Solar System structure (asteroids, NEOs, KBOs)
- Dark Matter, Dark Energy, Cosmology (spatial distribution of galaxies, gravitational lensing, SNe, AGN)
- Time Domain (cosmic explosions, variable stars)



LSST Alert Generation

- Compare each image with a reference via subtraction
- Report results within 60 seconds
- Some annotation and filtering within the Project
- Full stream(s) flow(s) out to community
- One to ten million alerts per night, every operational night, for ten years
- (One to ten thousand every 37 s)

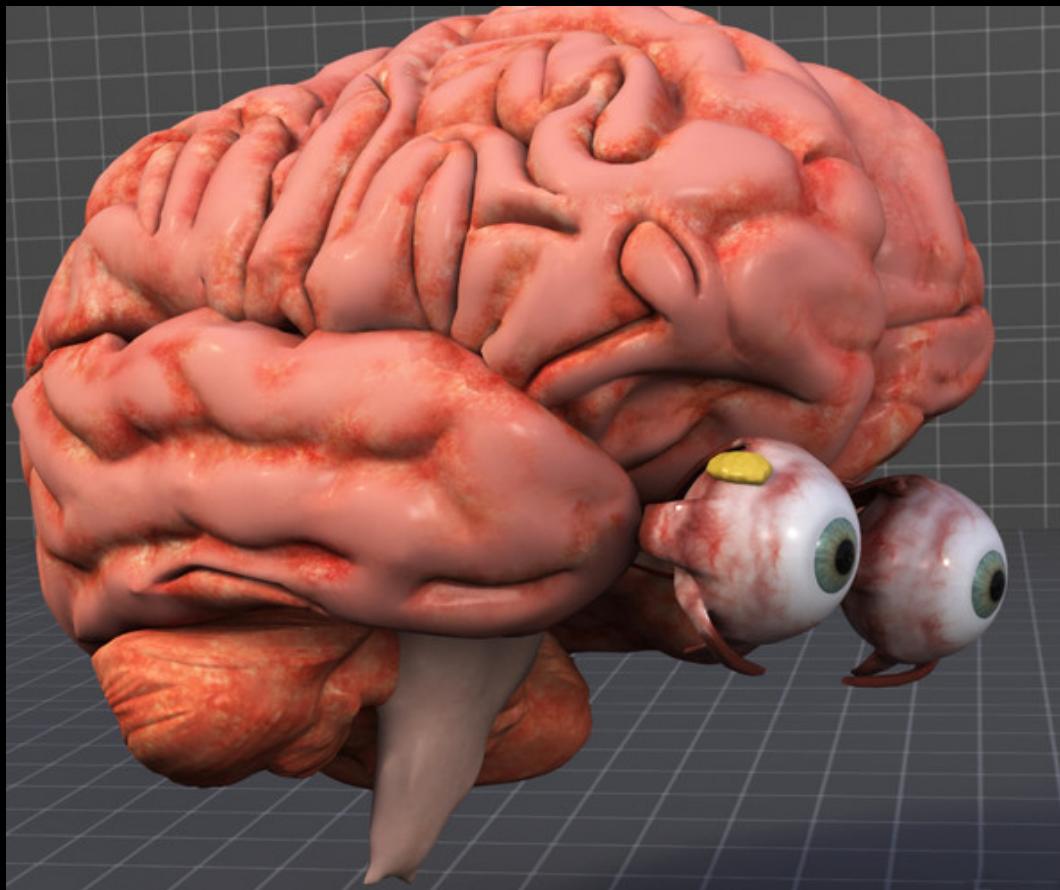






We need a broker, an intermediary who sits between the source of the alerts and the consumers of the alerts while filtering and adding value

Astronomical Brokers



Original Broker, prehistoric to ca. 2000
(minimum components)

Despite known
limitations with this
broker, has worked
extremely well

Volume and rate
exceeding capacity

Still necessary
Training sets
Pathological cases
EPO opportunities

Astronomical Brokers

Plenty of successful examples to build upon
A fundamentally multi-disciplinary problem

PUBLICATIONS OF THE ASTRONOMICAL SOCIETY OF THE PACIFIC, 124:1175–1196, 2012 November
© 2012. The Astronomical Society of the Pacific. All rights reserved. Printed in U.S.A.

Automating Discovery and Classification of Transients and Variable Stars in the Synoptic Survey Era

J. S. BLOOM,^{1,2} J. W. RICHARDS,^{1,3} P. E. NUGENT,^{1,2} R. M. QUIMBY,^{4,5} M. M. KASLIWAL,^{4,6} D. L. STARR,¹ D. POZNANSKI,^{1,2,7}
E. O. OFEK,^{4,8} S. B. CENKO,¹ N. R. BUTLER,^{1,9} S. R. KULKARNI,⁴ A. GAL-YAM,⁸ AND N. LAW¹⁰

Received 2011 June 27; accepted 2012 September 12; published 2012 October 24



● **SkyAlert**

To: Tom Matheson

SkyAlert event CBAT#876696 (71.874380, 23.982080)

SkyAlert email about event CBAT#876696

At 2013-11-07T08:02:32.64000Z, RA,Dec = (71.874380,23.982080)

The portfolio around this event is at <http://skyalert.org/events/876696>

The trigger observation alone is here: <http://skyalert.org/event/980597>

The XML for the trigger observation is here: <http://skyalert.org/event/xml/980597>

Your alert named " was the cause of this message with this trigger condition:

True

(A real-time action, such as this message, occurs when the trigger condition is true *because* of the trigger event, but is not true without it).

To cancel these alerts, go to <http://skyalert.org/rules/> and change/delete your alerts. You will have to login. Or just write help@skyalert.org

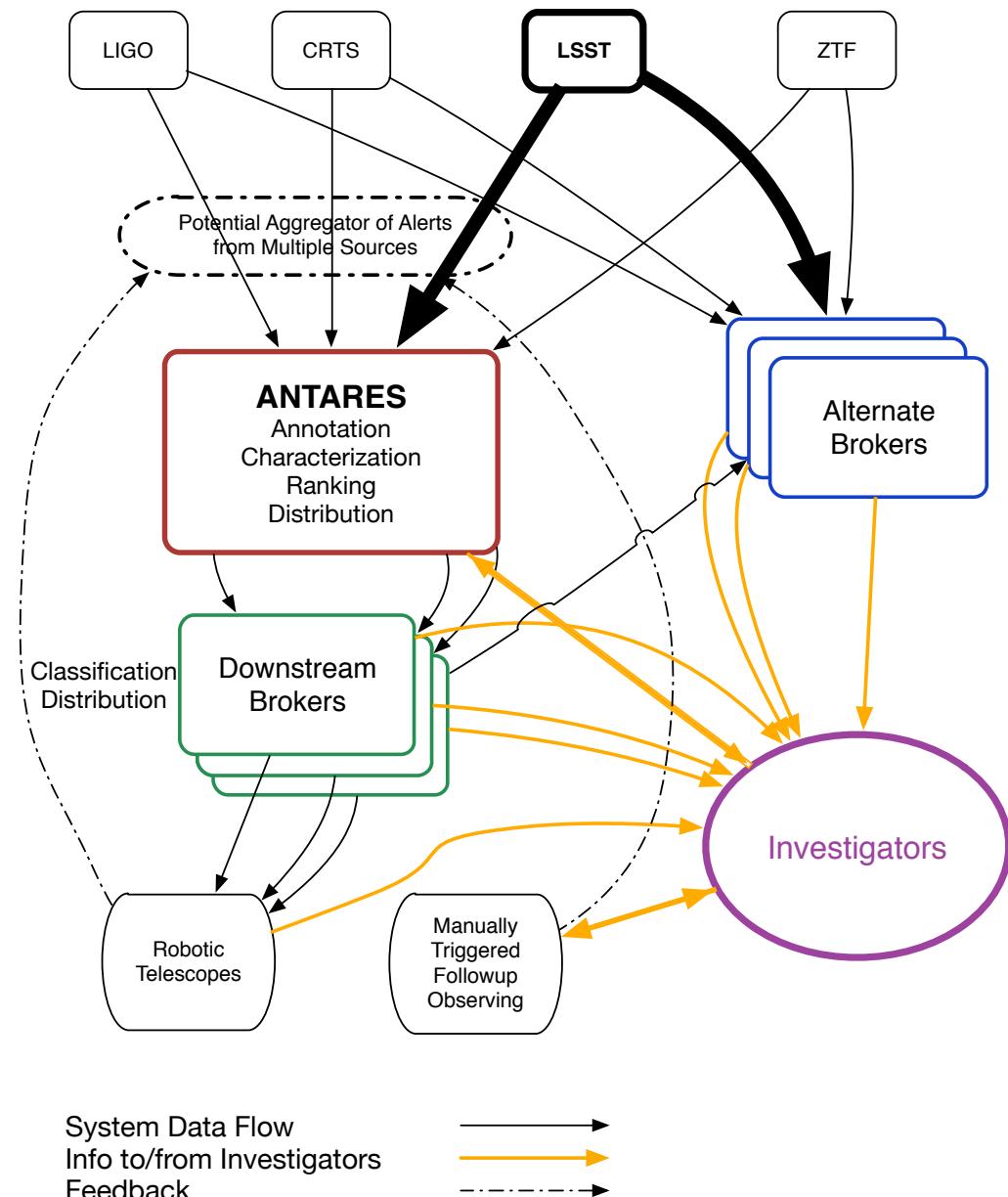


ANTARES: The Arizona-NOAO Temporal Analysis and Response to Events System

Funded by NSF INSPIRE proposal (CISE-AST-1344024)

ANTARES Environment

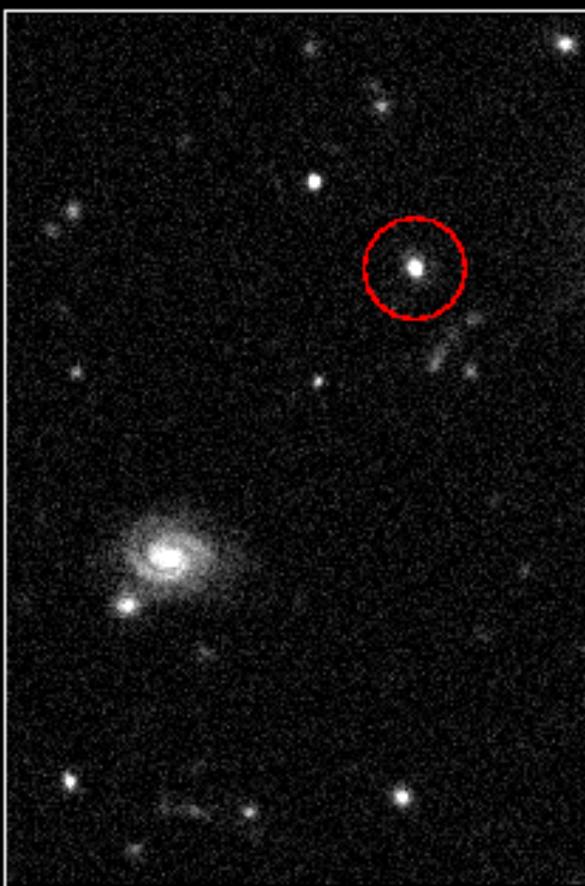
ALERT GENERATORS: Difference Imaging, Real/Bogus & Moving Object Assessment



How ANTARES Fits into the Overall Astronomical Ecosystem

What it does, and does not, intend to do

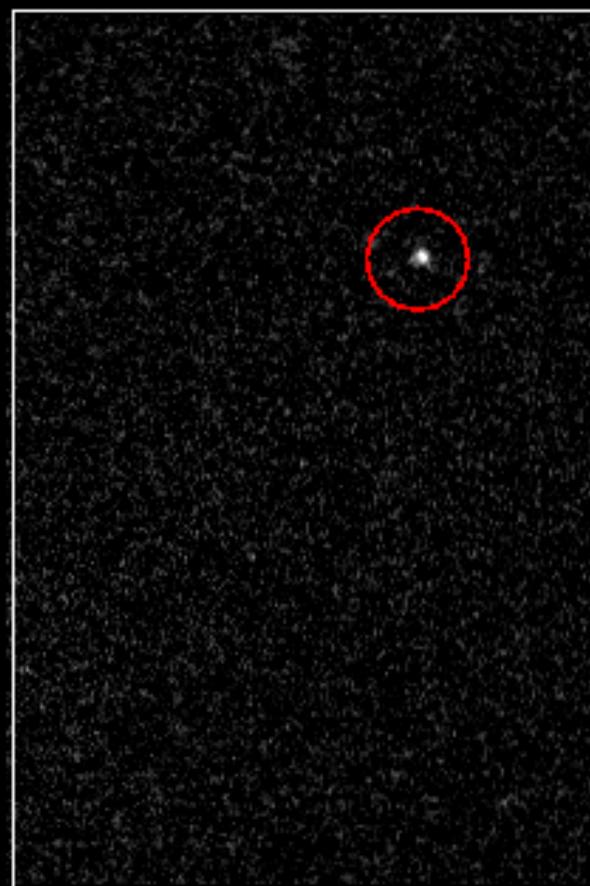
Epoch 1

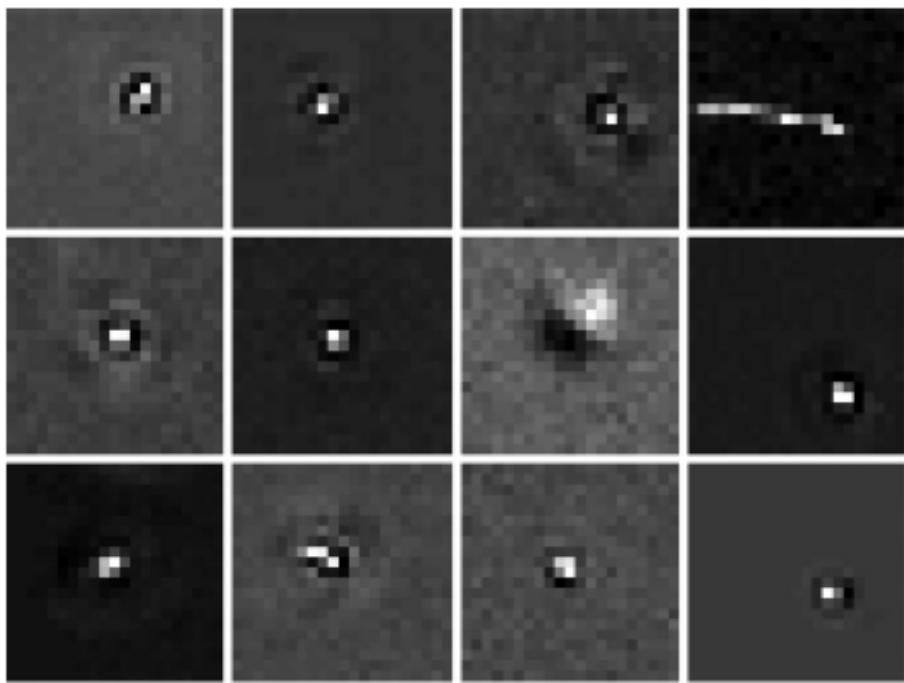


Epoch 2

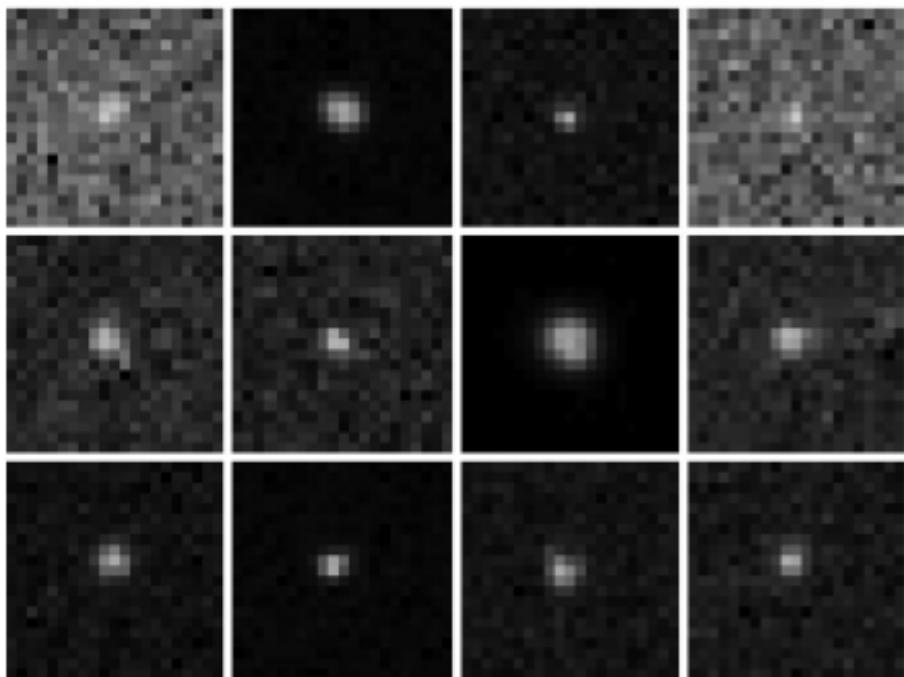


Epoch 2 - Epoch 1





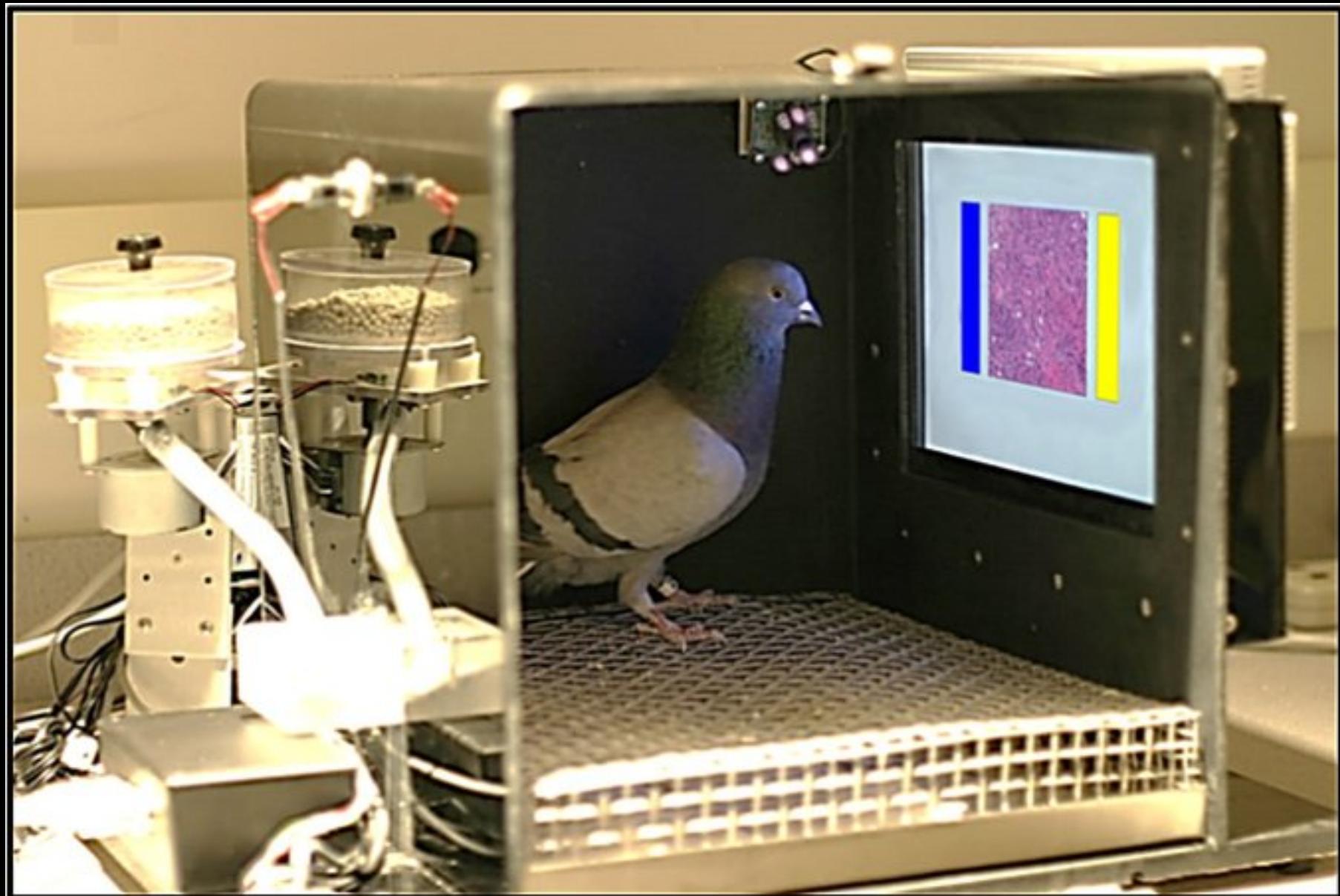
Subtractions can produce many false alerts. Distinguishing them is a job for the people with deep knowledge of the camera system.



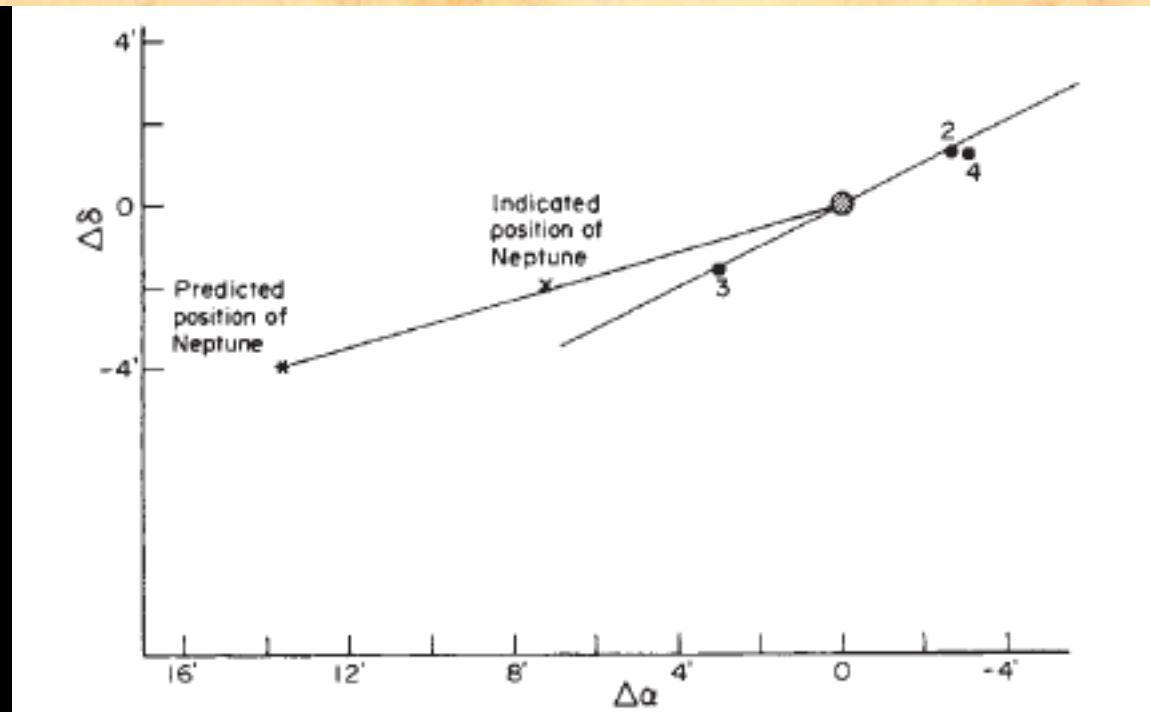
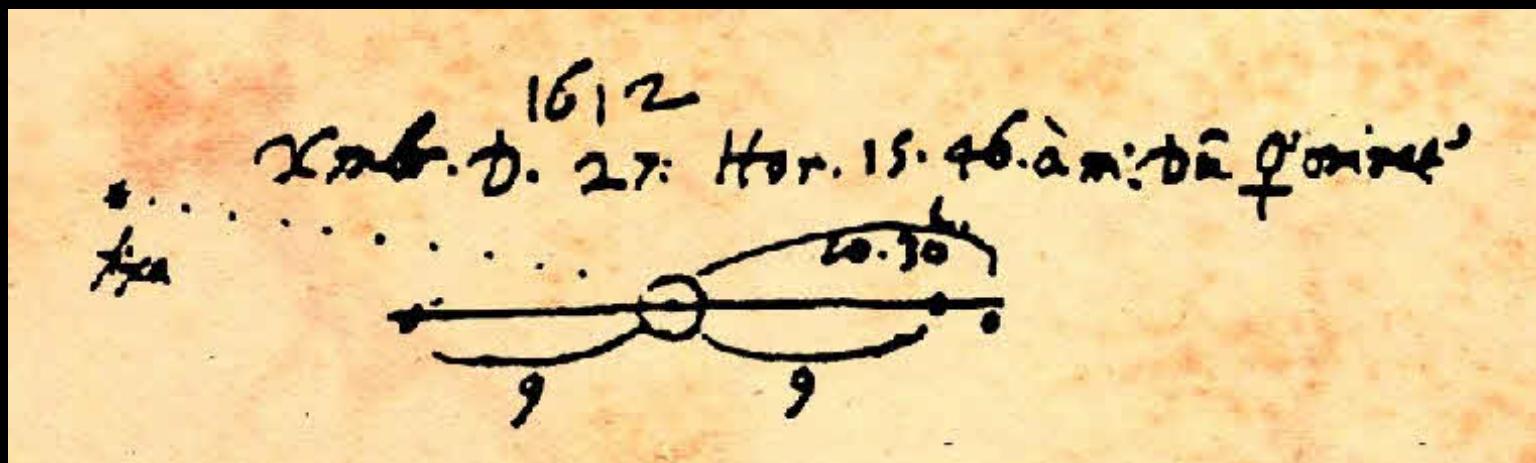
Josh Bloom calls this real/bogus detection

LSST will do this filtering and report a ‘bogosity’ for each alert, something we can use to filter

We still end up with a few million alerts



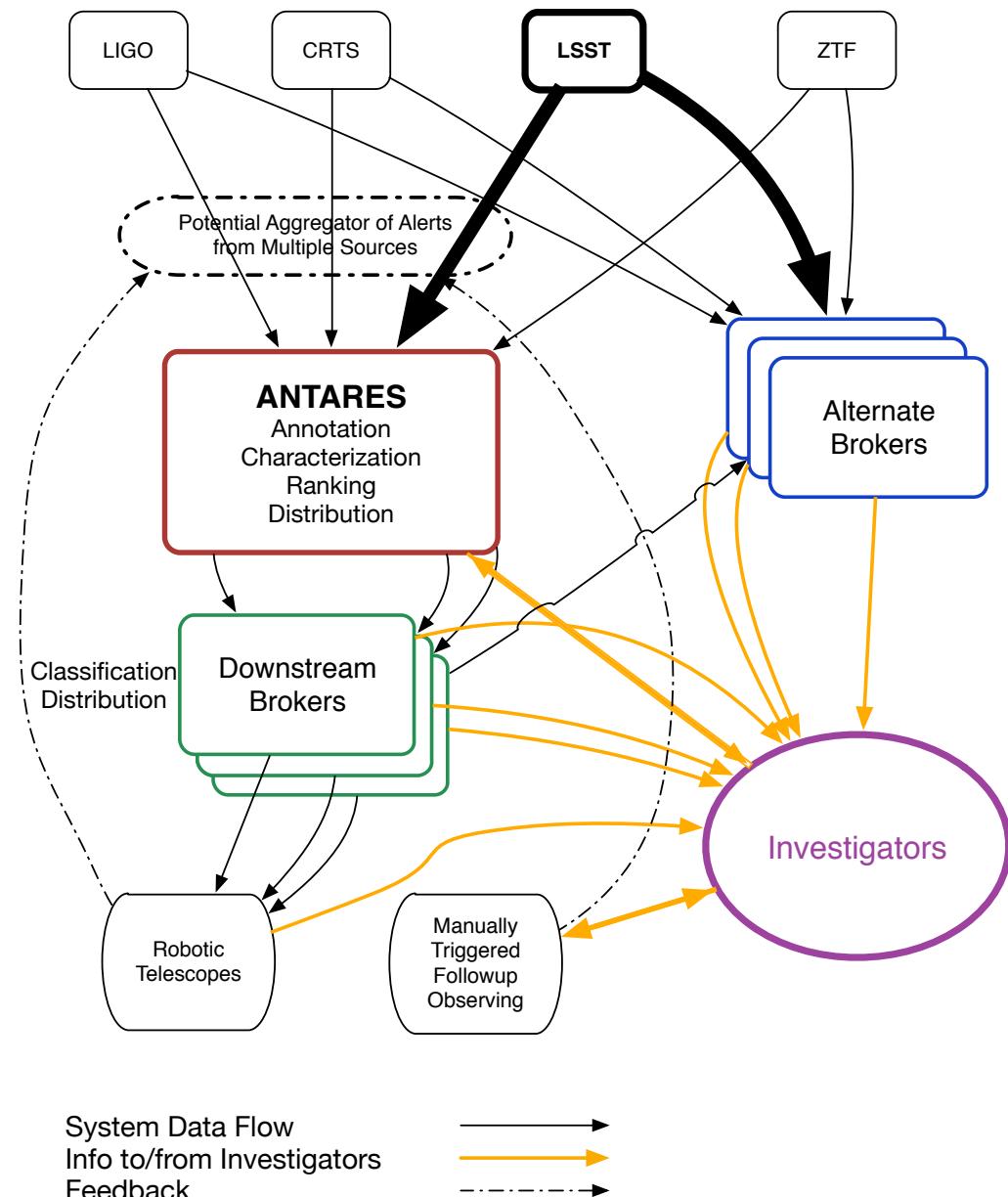
What Does Discovery Mean?



Kowal & Drake
1980

ANTARES Environment

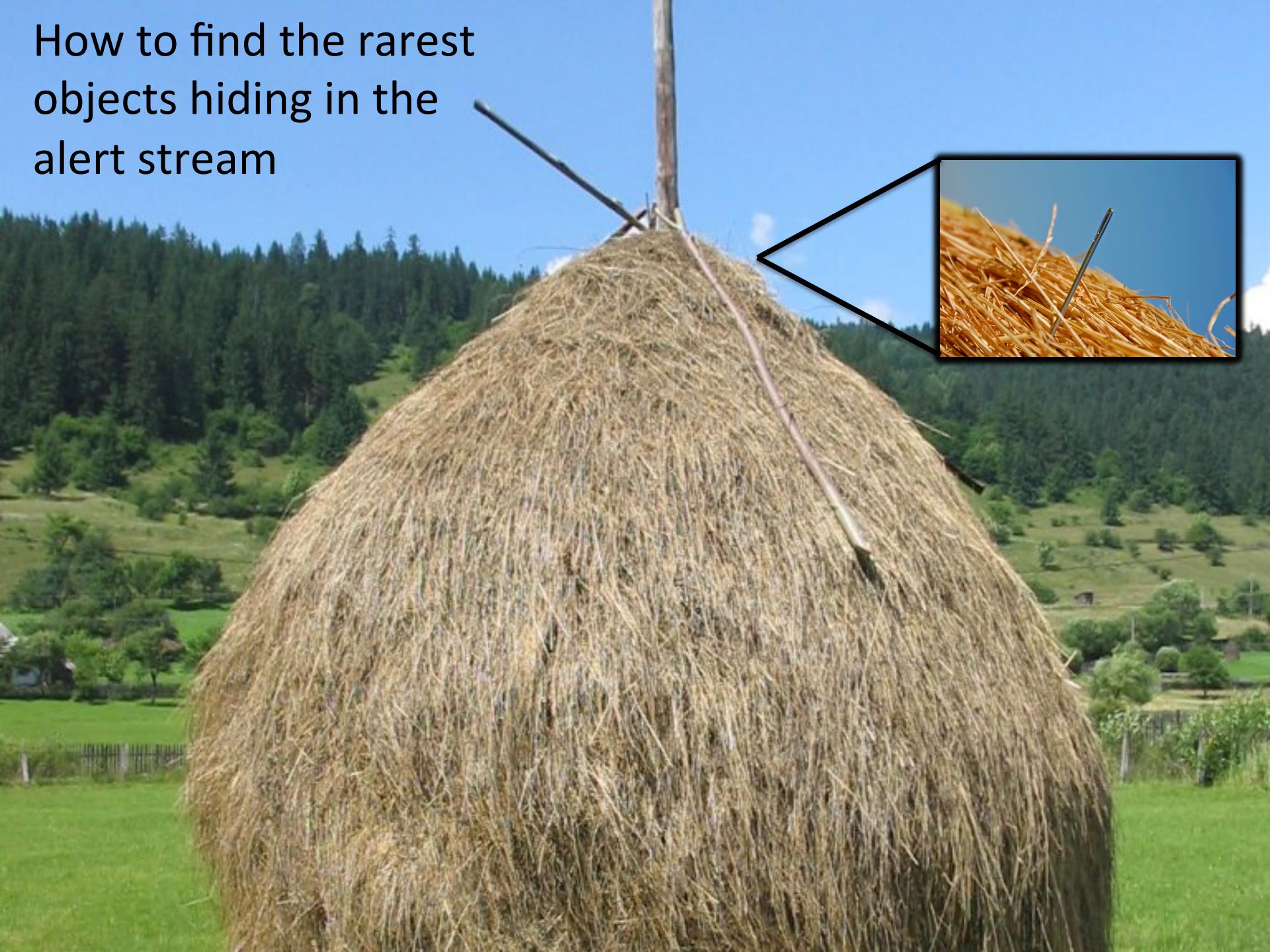
ALERT GENERATORS: Difference Imaging, Real/Bogus & Moving Object Assessment



How ANTARES Fits into the Overall Astronomical Ecosystem

What it does, and does not, intend to do

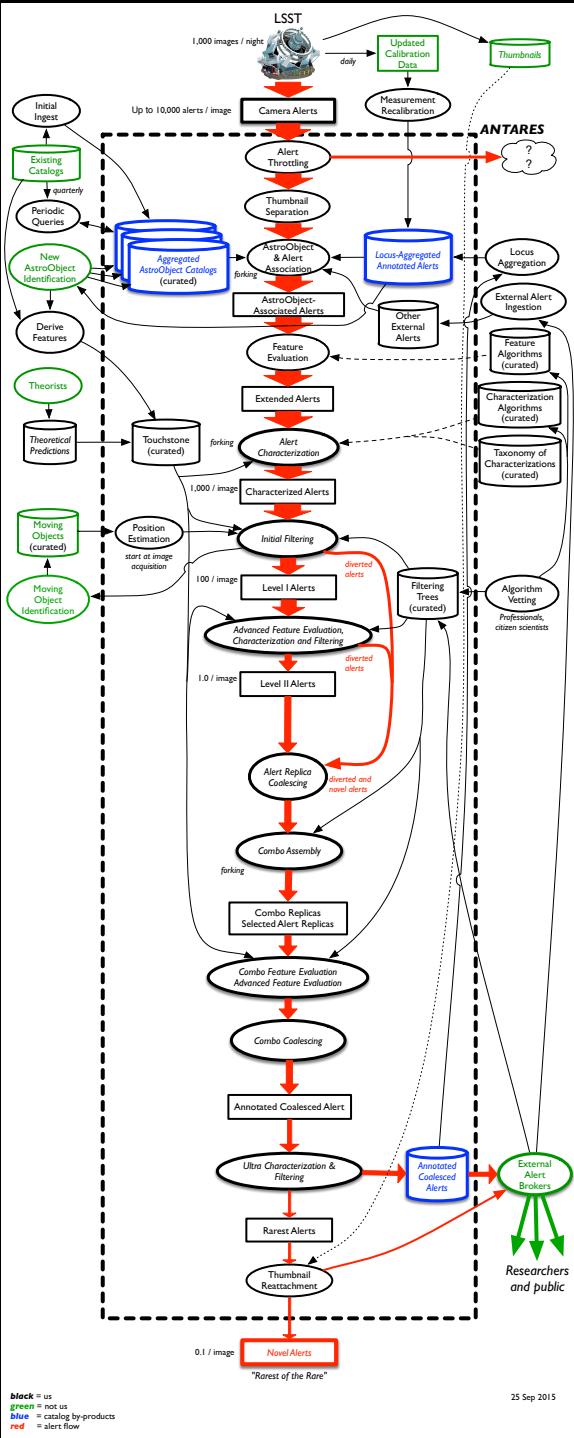
How to find the rarest objects hiding in the alert stream



Given the limited information in each alert,
this is the real problem



ANTARES Architecture



Open source/open access

Community driven filters

The prototype is focused on finding the 'rarest of the rare'

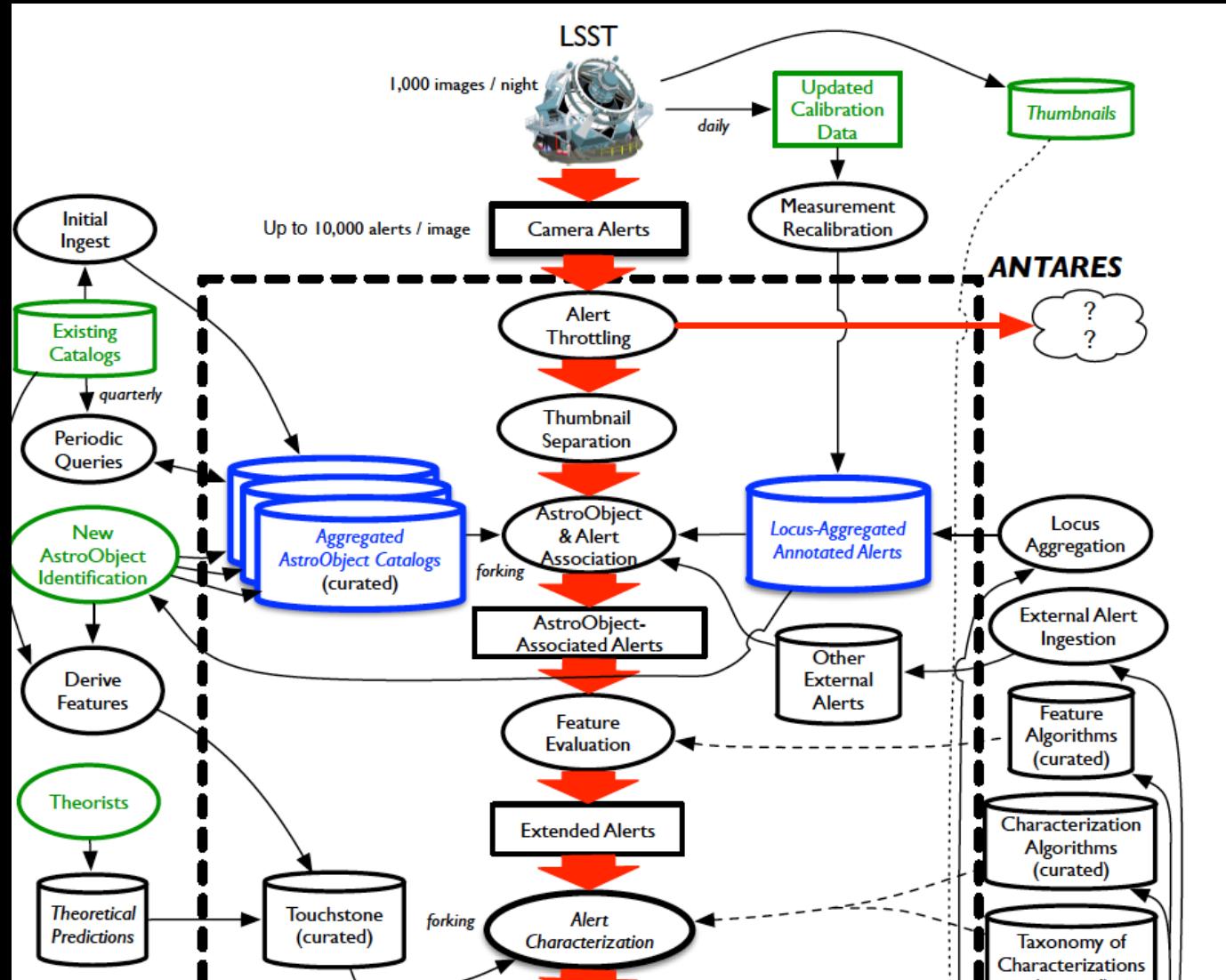
This tests the core flow of the system

Nothing is lost; we don't throw out any alerts

Future versions could accommodate multiple filtering paths to address many goals

Volume and rate are key problems

ANTARES: Annotation and Added Value



Source association is a critical step, positional

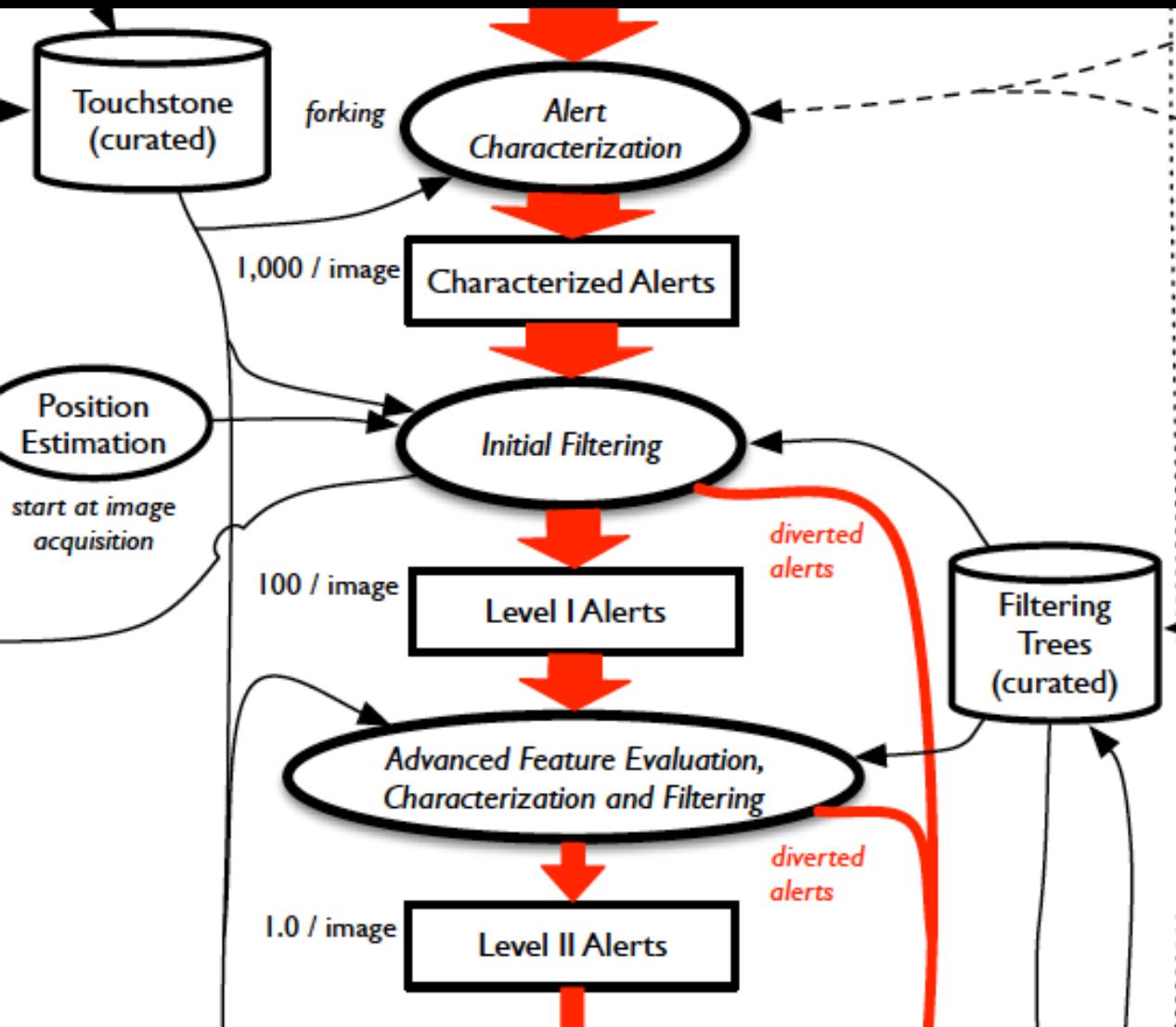
Annotation of alerts with external information

e.g., Sloan, 2MASS, Chandra; eventually LSST

The annotated alerts are available to outside brokers

Derive features from alert+annotation+history

ANTARES: Filtering Algorithms



Multi-step process

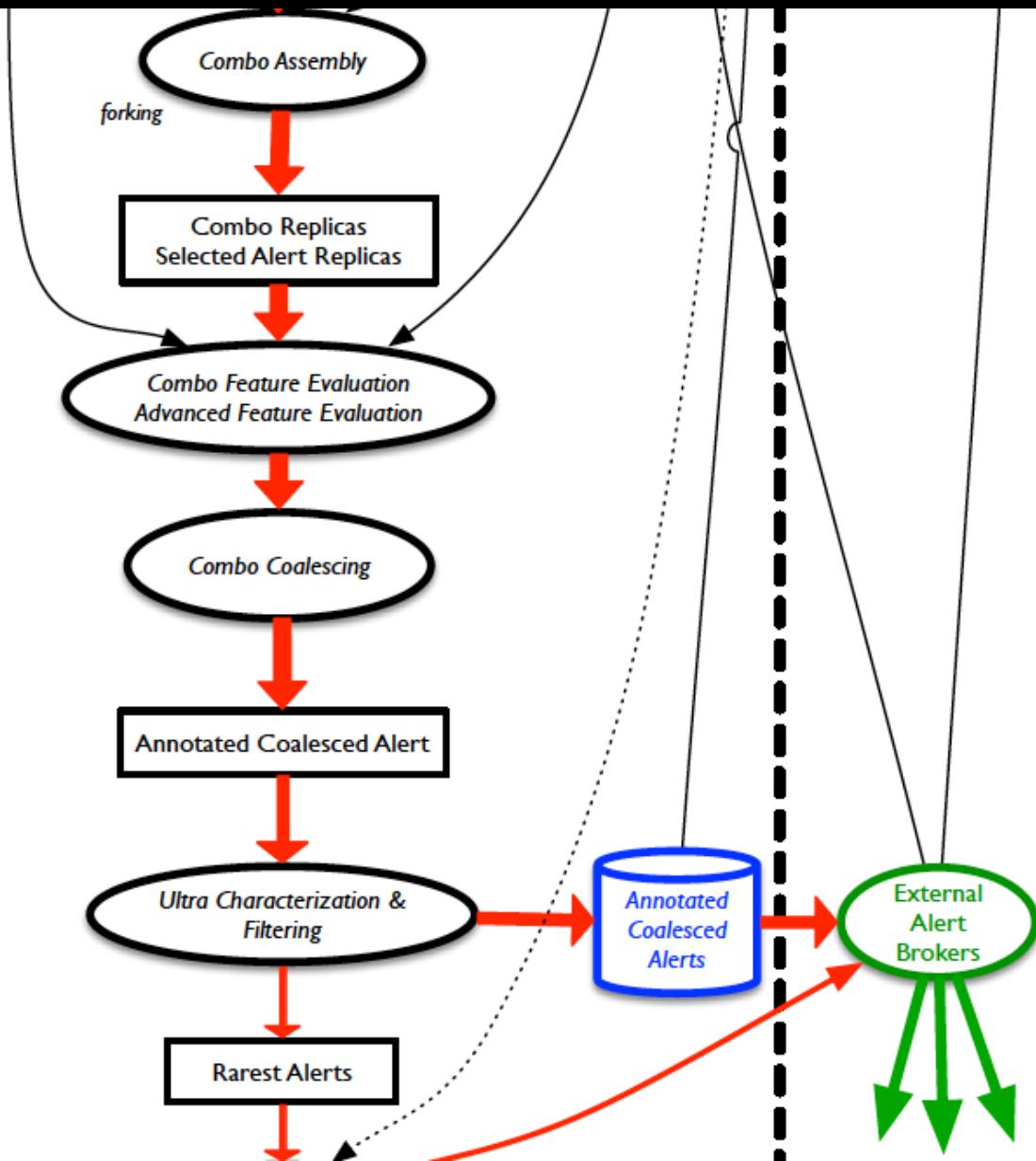
'Categorization' allows for winnowing of alert stream

Algorithm expertise will help with rapid processing, but also filter order/scheduling

More complex filters only applied to subset of alerts

User-configurable filters and algorithms

ANTARES: Filtering Algorithms



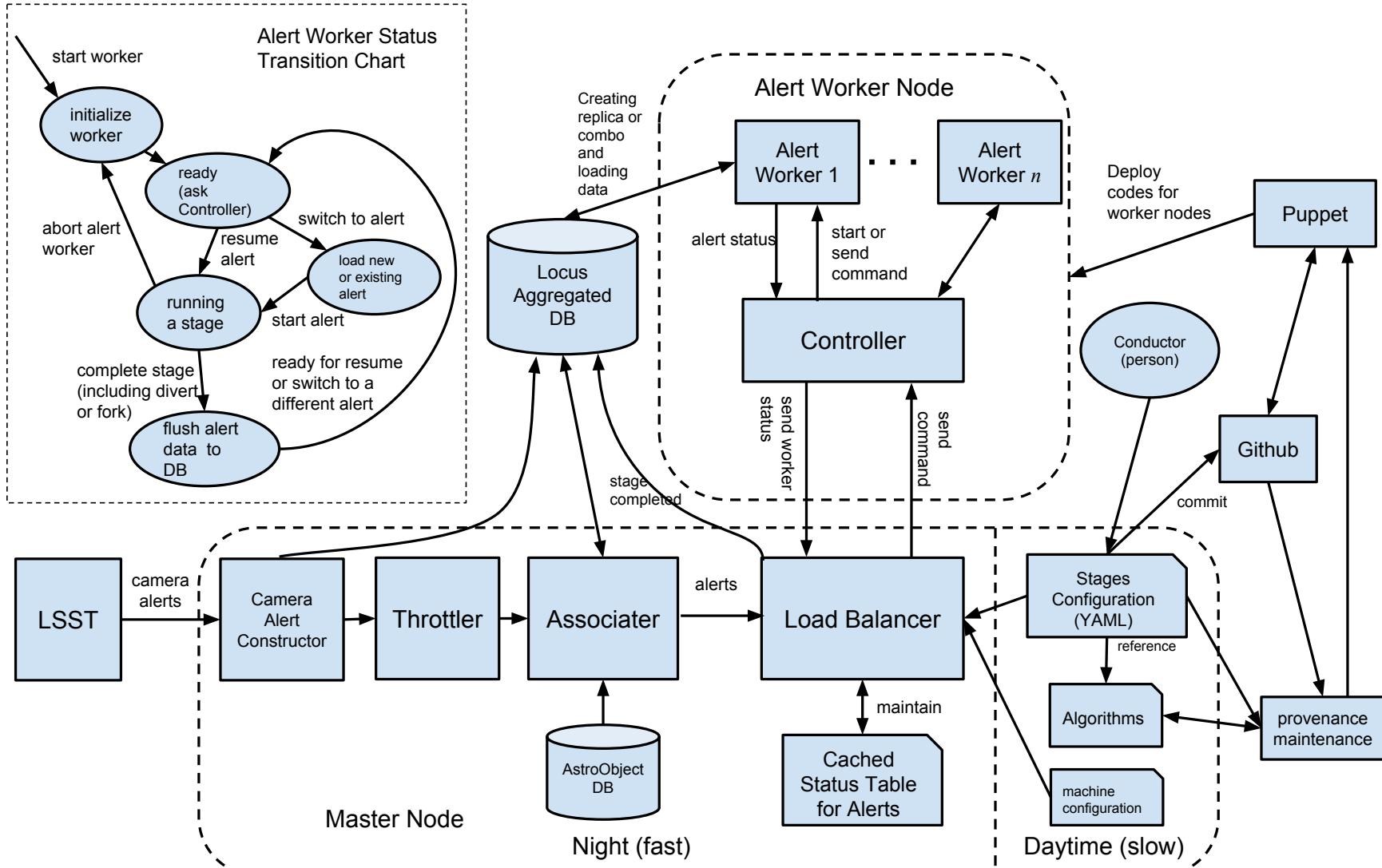
'Combos' consider multi-object associations

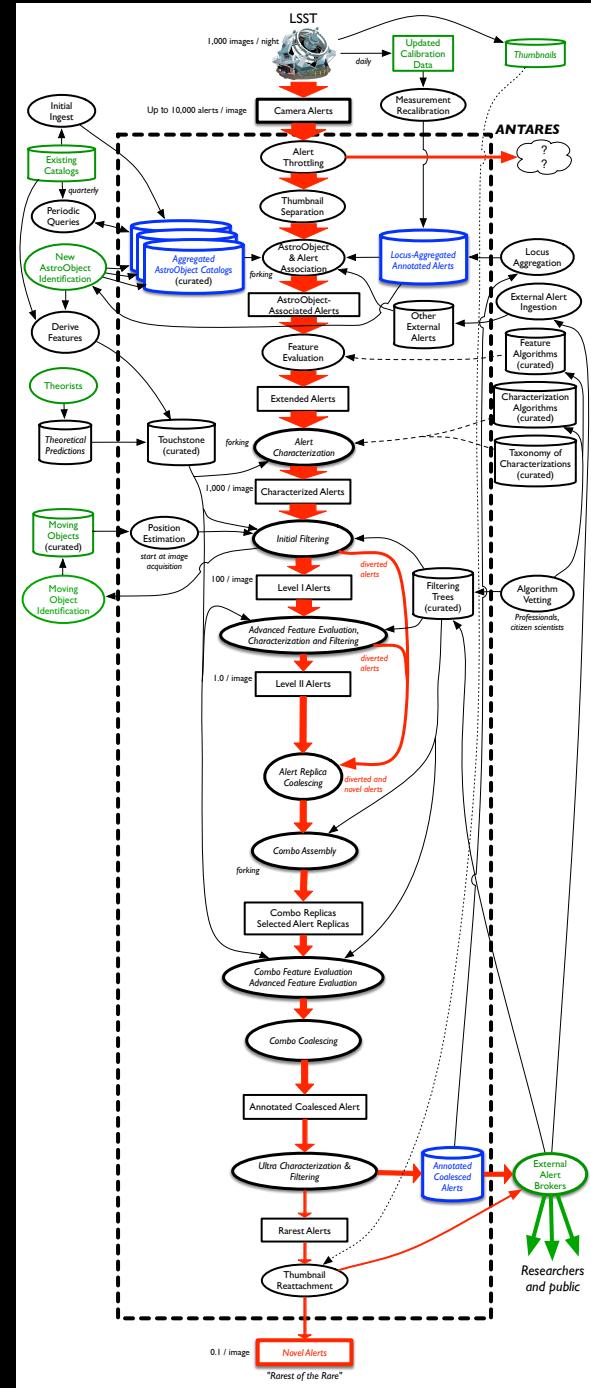
As the volume of alerts decreases, processing complexity increases

This focuses on rare and short-lived objects. Downstream brokers (which could be ANTARES copies) can look through the annotated alert database on a range of timescales.

Fully Parallel Load Balancing Architecture with Provenance

Antares Load Balancing Diagram





black = us
green = not us
blue = catalog by-products
red = alert flow

What ANTARES is Doing Now

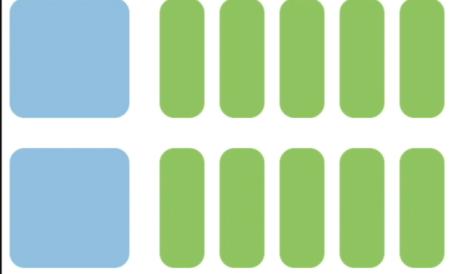
192.168.56.5:5000

ANTARES Dashboard

All information Final Decision System Warnings

Start Monitoring

Send Alerts

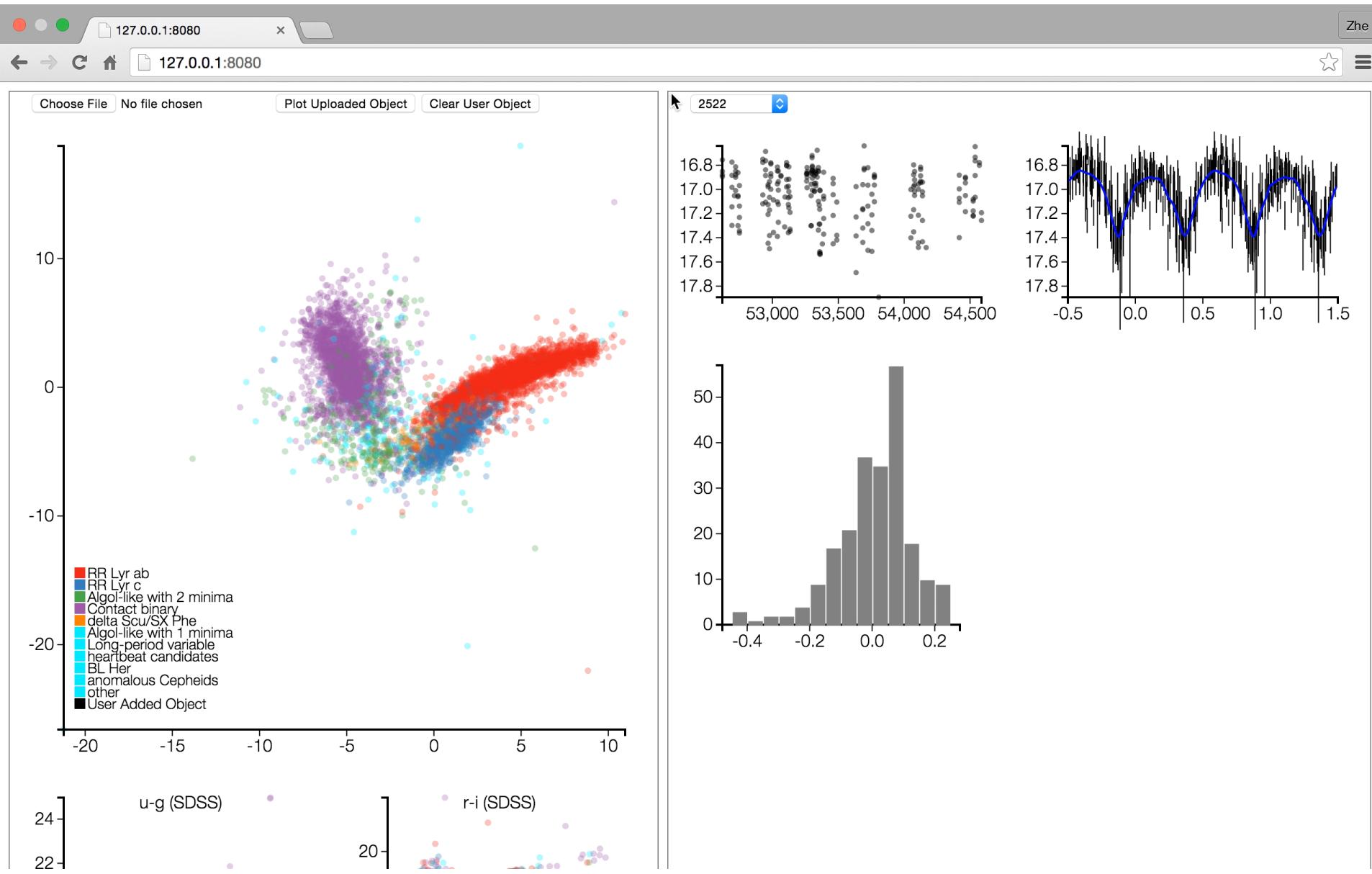


Two sets of five vertical bars each, colored blue and green. The top set has the first bar blue and the next four green. The bottom set has the first bar blue and the next four green.

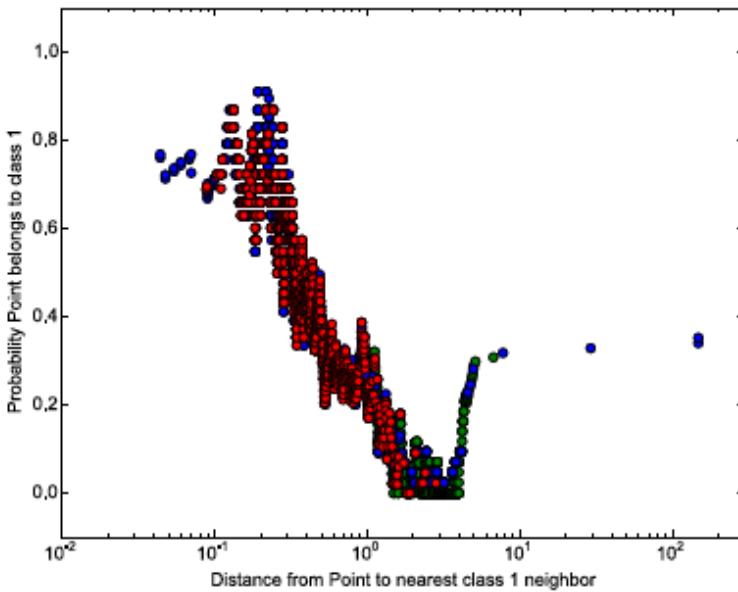
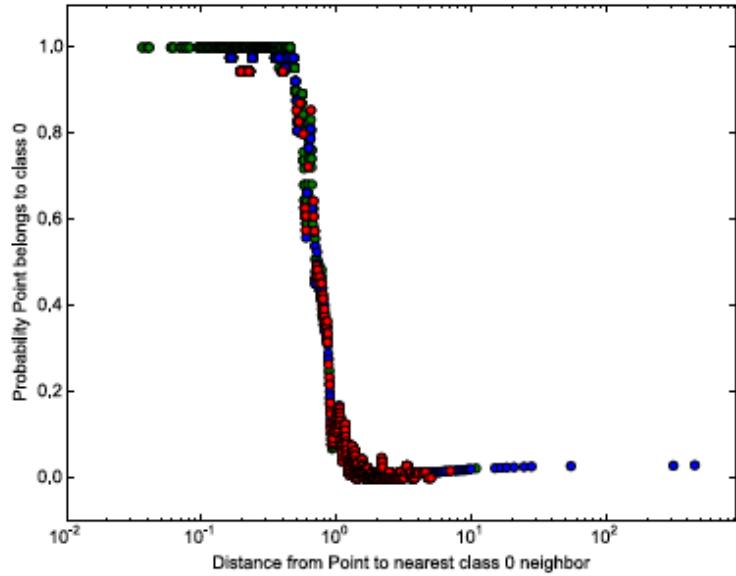


A large empty rectangular area for dashboard content, likely a placeholder for real-time data or graphs.

Visualization to Explore Feature Space

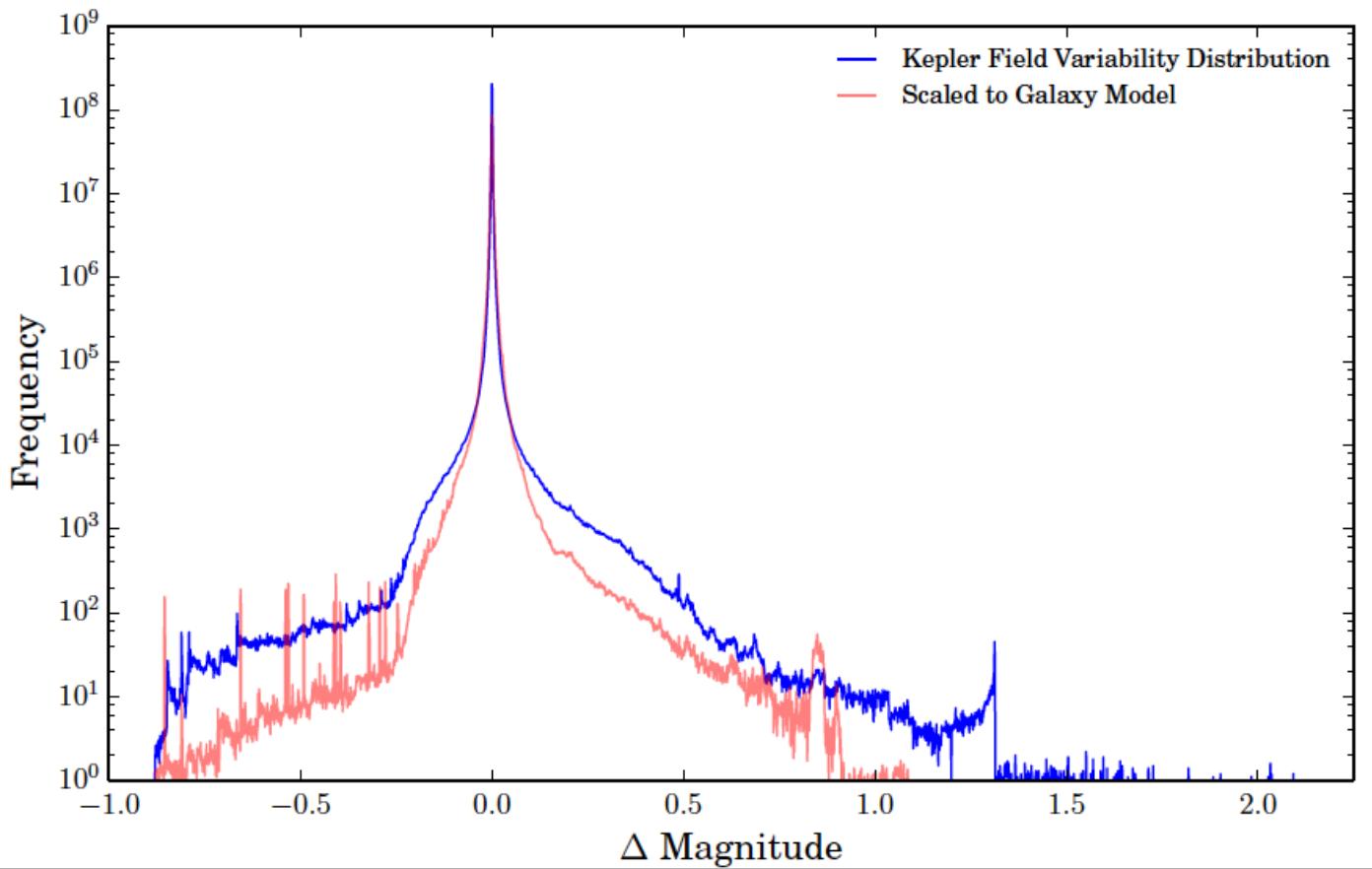


Algorithms for Feature Space Comparison



- Lots of potential features: light curve parameters, location on the sky, nearest object, etc.
- Fundamentally high dimensional
- Touchstone attempts to distill astronomical knowledge
- Many algorithms possible (kNN, random forest, etc.)
- Our modified kNN is fast and accurate
- Can use other techniques to reduce dimensionality (e.g., PCA)

How to Make a Decision When you Know Nothing

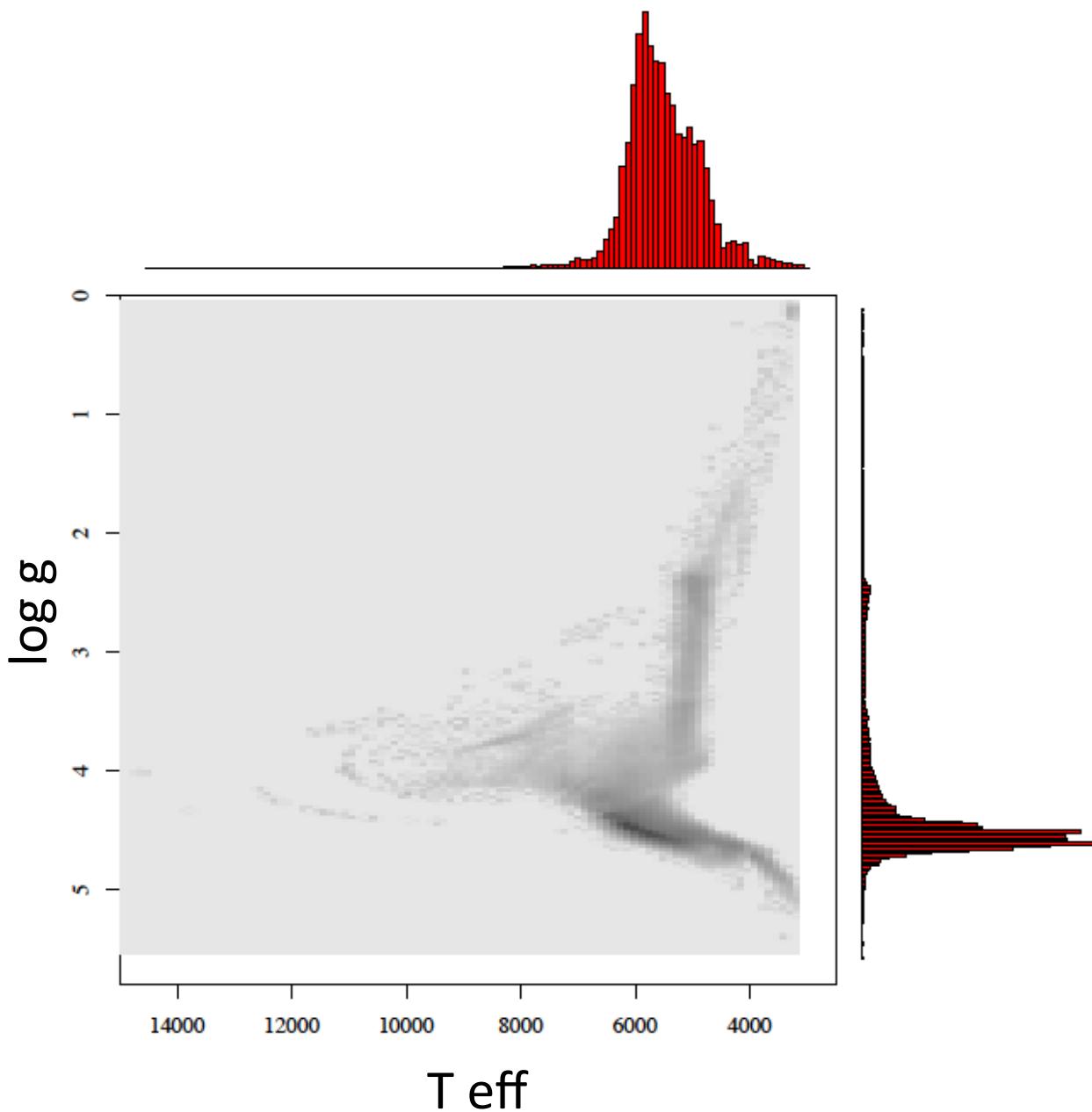


Use Kepler photometry to establish baseline variability distribution

Use Besançon Galaxy models to predict stellar distribution

Warp Kepler variability to match new stellar distribution

How to Make a Decision When you Know Nothing

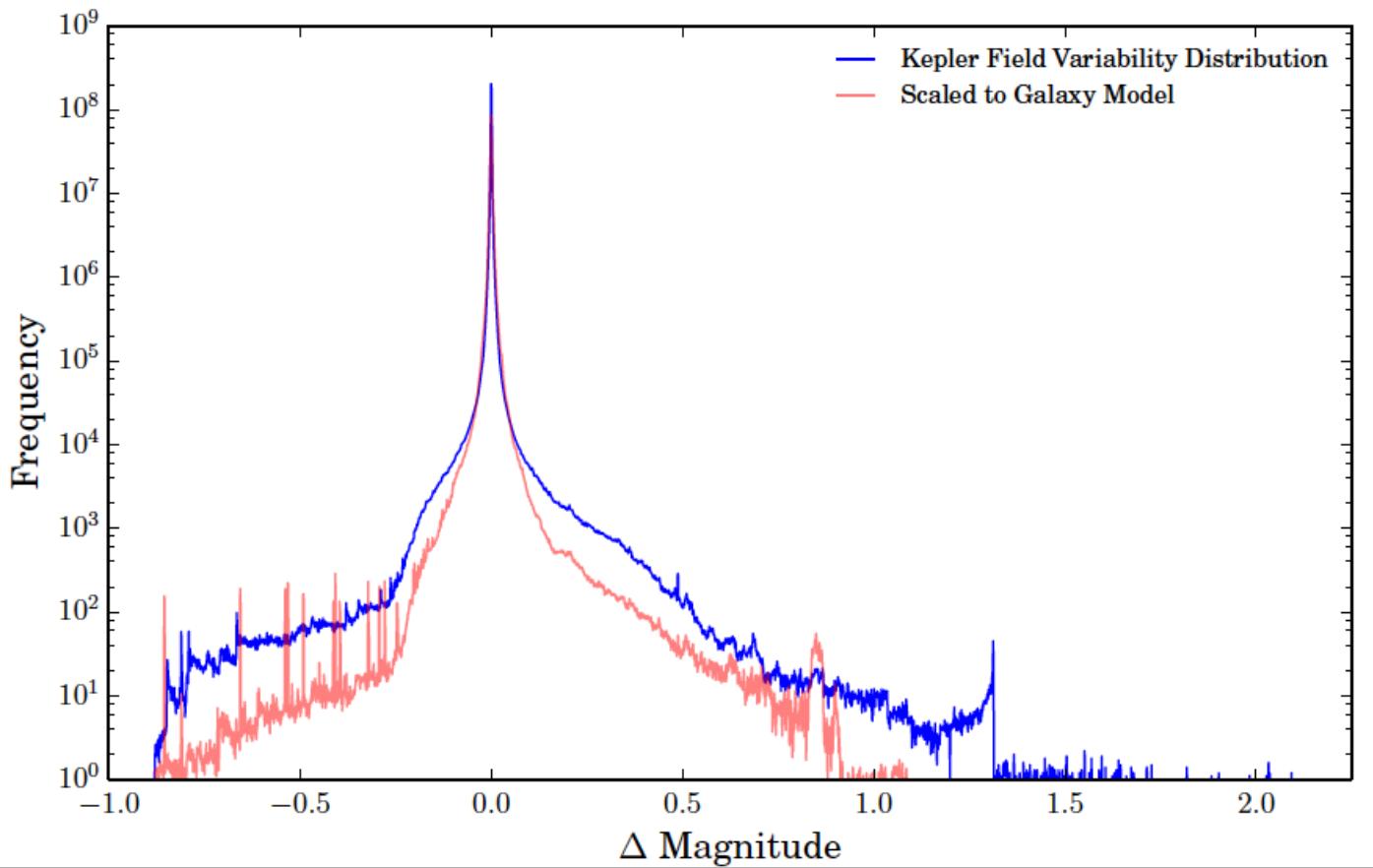


Use Kepler photometry to establish baseline variability distribution

Use Besançon Galaxy models to predict stellar distribution

Warp Kepler variability to match new stellar distribution

How to Make a Decision When you Know Nothing



Use Kepler photometry to establish baseline variability distribution

Use Besançon Galaxy models to predict stellar distribution

Warp Kepler variability to match new stellar distribution

Where Do We Go from Here?

- More external catalogs
- More features
- More feature-space comparison
- Scaling and optimization
- Sophisticated statistical techniques
- Fast database design
- Provenance control
- Live event streams

Where Do We Go from Here?

- We need your help
- Domain expertise is the best way to develop features that distinguish objects
- Labeled time-domain data establishes a baseline feature set
- Go see Flash talks this Friday, Gautham Narayan and Zhe Wang will show even more