LSST community broker call: A French initiative

Emille Ishida, Julien Peloton for the LSST-France Broker Initiative

The role of community brokers

What the brokers must do:

- Digest the alert stream
- Add value to alerts: *enrich each alert with extra information, either from connection with existing catalogs or from preliminary ML classifications*
- Distribute alerts to the community

What the brokers can do:

- Adapt the filtering algorithm to their interest as the survey evolves
- Coordinate spectroscopic follow-up allocation for a particular science case
- Decide what to distribute and what to keep private

What users most expect from brokers? *(according to the survey conducted by TVS)*

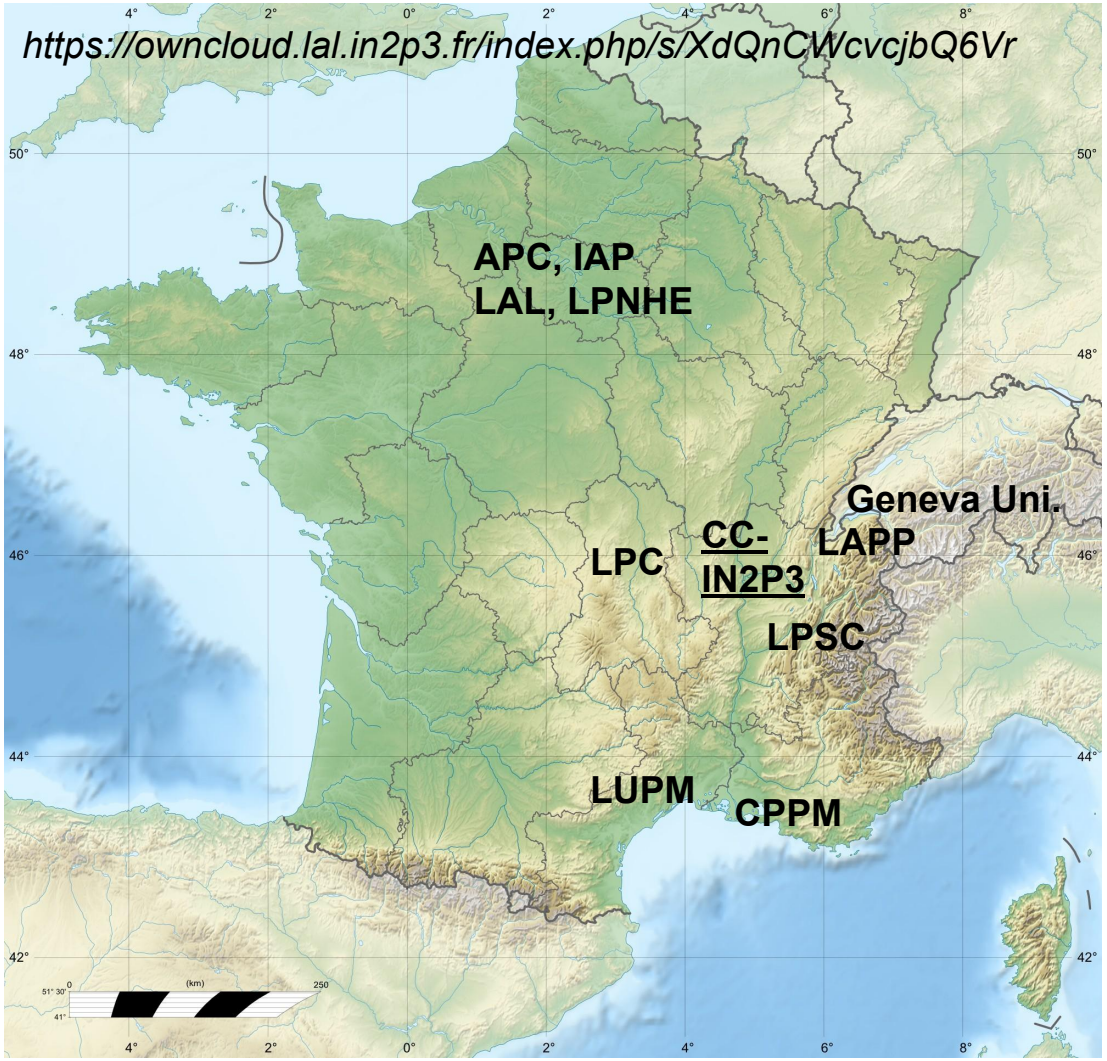
- Early classification of identified sources
- Cross-match with alert streams and data bases in other wavelengths (MMA)

LSST alert stream will certainly be a mine for astronomical discovery!

Whoever controls the a broker will have a large advantage in identifying and exploiting it.

Letter of Intent

A letter of Intent has been submitted to LSST DM to propose a community broker
Only a **few brokers** will have access to the LSST data stream.



Initiative mainly from IN2P3 labs*.

31 endorsers: ⅔ LSST-France, ⅓ outside.

Focus on: **Supernovae**, **micro-lensing**, and **multi-messenger astronomy**

Links with: Fermi-LAT, DESI, CTA, Euclid, Virgo, ENGRAVE, SVOM, INTEGRAL, ANTARES/KM3NeT.

** Will be opened to anyone in a second time*

Scientific goals

Optimize the scientific impact of LSST alerts for a set of science cases

Supernovae

Time Frame: days to months

- ≈30.000 well sampled type Ia SNe light curves per year
- ≈1000 are expected spectroscopically confirmed per year

Microlensing

Time Frame: hours to months

- Large impact on extra-solar planet search
- Explore the mass distributions and the Milky Way dynamics.
- LSST will reach fainter magnitudes than OGLE (23 vs 20 mag)

Multi-msg Astronomy

Time frame: minutes to days

- Optical counterparts for many experiments: CTA, LIGO, Virgo, KAGRA, SVOM, KM3NET, …

Anomaly detection

Time Frame: minutes to months

- Unforeseen astronomical sources is one of the most exciting outcomes.
- Should be more pronounced after a few seasons.

Photometric SN cosmology

CPPM LPC LPNHE LUPM +

Supernovae

Time Frame: days to months

- ≈30.000 well sampled type Ia SNe light curves per year
- ≈1000 are expected spectroscopically confirmed per year

Unique expertise in France.

Years in working with spectroscopic follow-up programs.

Ideally spectra should be taken around maximum light.

Must optimize spectroscopic resources to ensure optimal training for machine learning applications.

Microlensing

LAL IAP +

Microlensing

Time Frame: hours to months

- Large impact on extra-solar planet search
- Explore the mass distributions and the Milky Way dynamics.
- LSST will reach fainter magnitudes than OGLE (23 vs 20 mag)

Some team members are precursors of the microlensing science (EROS).

Microlensing can probe to great distances from the Sun (>8.5 kpc)

Milky Way: expected to detect an average rate of 15 events/deg2/year in the disk and 400 events/deg2/year in the Bulge.

Link to networks like MicroFUN, or astrometric and spectroscopic follow-up observations.

Multi-messenger astronomy

APC LAL LAPP LPSC LUPM +

Multi-msg Astronomy

Time frame: minutes to days

- Optical counterparts for many experiments: CTA, LIGO, Virgo, KAGRA, SVOM, KM3NET, ...

New window opened recently with GW + neutrinos and gamma ray bursts.

Large increase of data flux at the 2020 horizon, and beyond: the challenge is to digest the very large number of alerts that will be sent.

LSST will provide a unique information in the visible band.

APC is developing an interface which manages the flow of VOevent alerts, and LAL a multi-messenger broker for the SVOM experiment (GRANDMA project)

Anomaly detection

LPC +

Anomaly detection

Time Frame: minutes to months

- Unforeseen astronomical sources is one of the most exciting outcomes.
- Should be more pronounced after a few seasons.

LSST is no more the place for serendipitous discoveries.

Algorithms should be adapted to return increasingly lower rate of ordinary objects with high anomaly score as the survey evolves.

Uses input from the user in order to fine tune anomaly scores.

Adaptive Machine Learning

- Early classification is the most pressing added values expected from brokers
- AL algorithms **enable** the construction of optimal training sets and **provide** classification and anomaly scores whose accuracy improves with the evolution of the survey.
- This paradigm can be couple to many different classifiers

We can deliver a broker with evolving added values which can optimize not only classification, but also discovery!

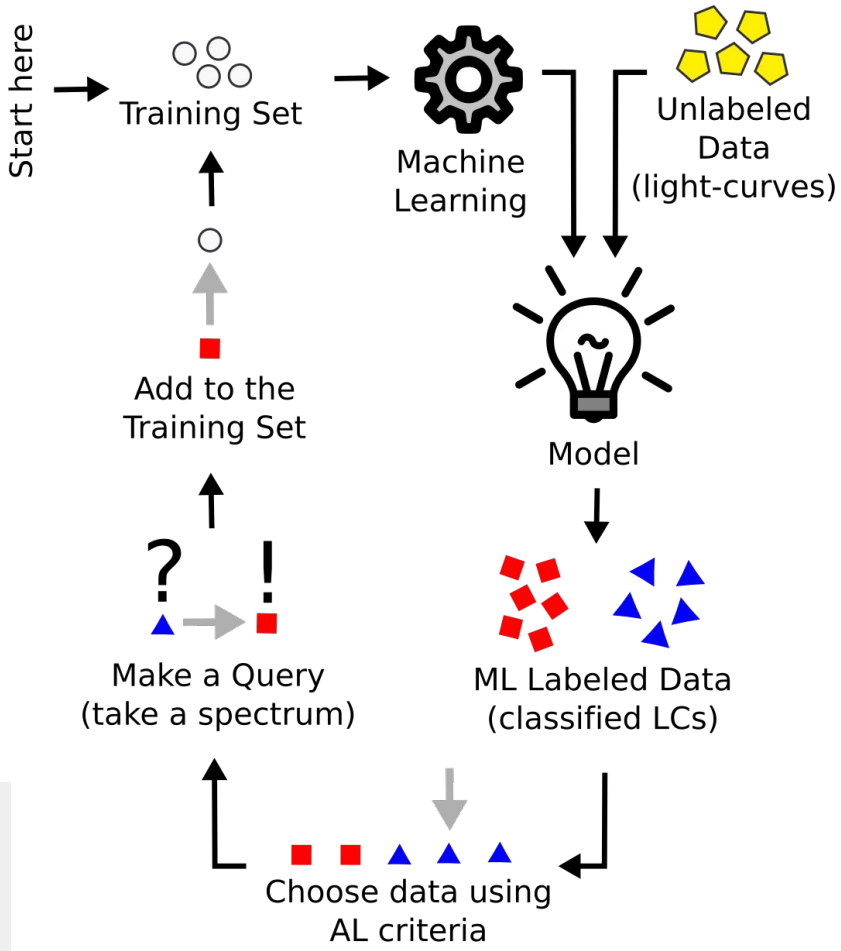
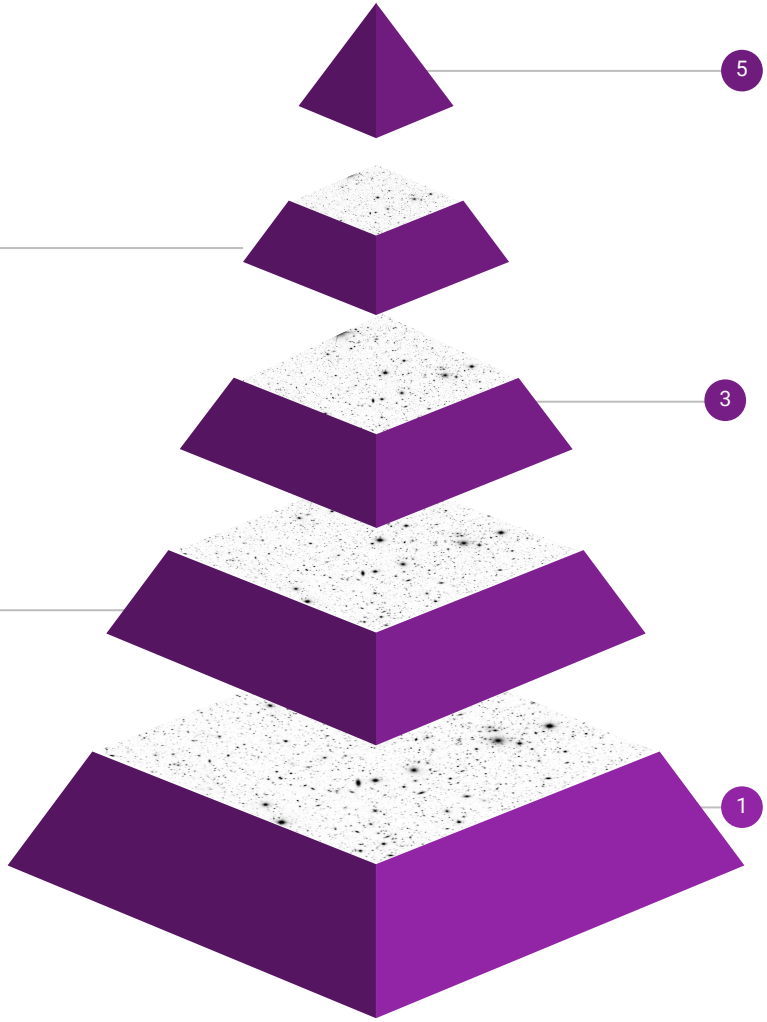


Diagram from: *Ishida et al., 2019, MNRAS, 483 (1), 2–18*

The LSST alert stream



Confirmed sources
5 New source!

Spectroscopic Observations
4 Organise follow-up

Cross-Match
3 Correlate to other experiments and catalogs

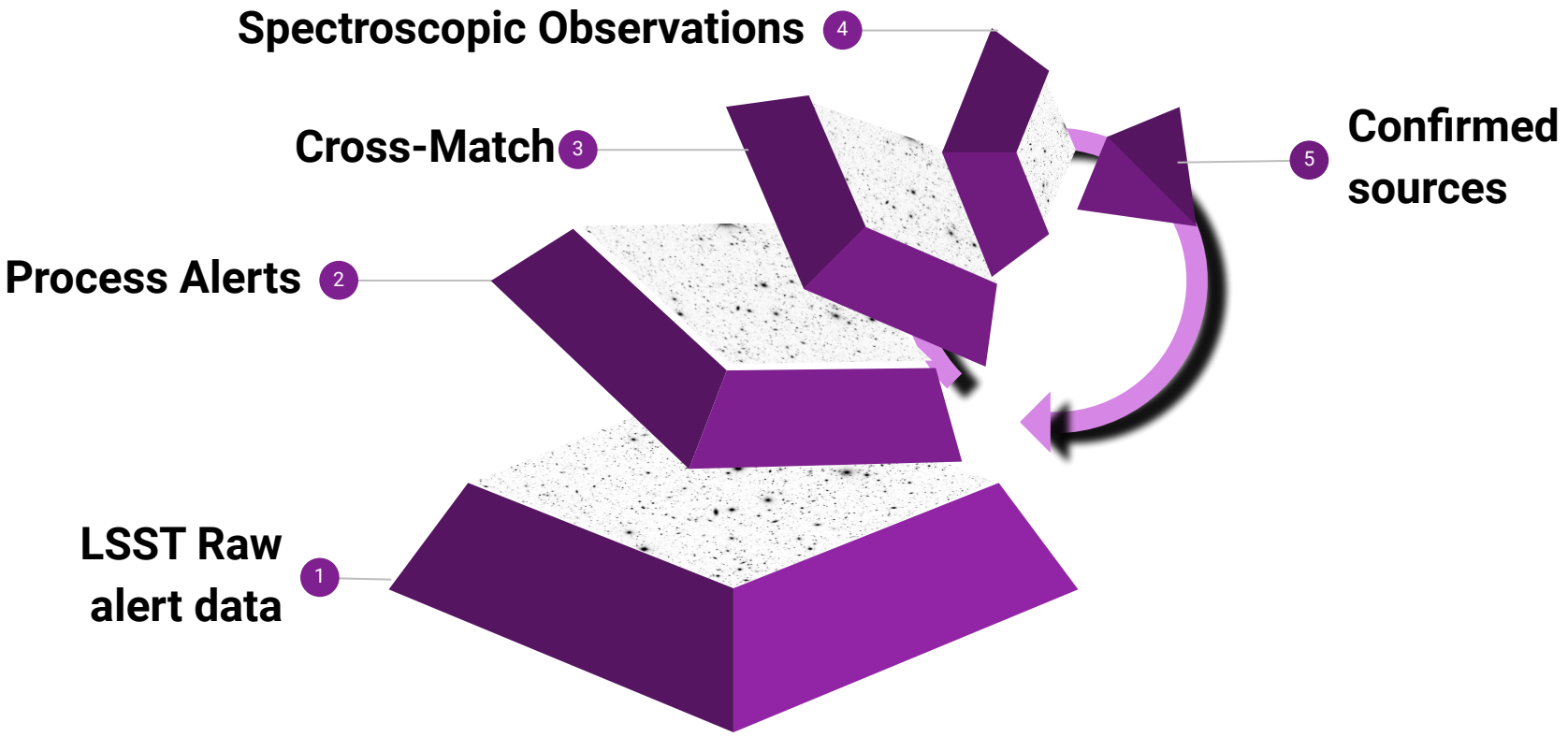
Process Alerts
2 Broker core algorithms

1 **LSST raw alert data**
Difference image analysis
~1TB/night expected.

Typical journey of an alert with Active Learning

The LSST alert stream

For classification: first years are crucial
For identification of anomalous sources:
things start to get interesting after bulk of data is categorized (a few seasons)



Deliverables

Continuously updated classification and anomaly scores by interactively identifying optimal training candidates for machine learning applications

1. Cross-match with existing catalogs and other experiment alert streams
2. Output from ML-based algorithms.
3. List of targets which should be prioritized by spectroscopic follow-up

Archive

Connect to large streams of alerts, and build databases.

Redistribute

Let other users and telescopes know about events of interest.

Process

Include broker services, user software programs, and science portal interface

Coordination

- Identifying interesting alerts is only part of the story: we need coordination with spectroscopic follow-up resources.

- At the end of each cycle we will publicize a prioritized list of targets that should be spectroscopically followed in order to improve future estimates.

- Workshop in Clermont Ferrand, 11-12 July:
  Optimizing training samples for SN photometric classification

- Must study deeply the possibility to coordinate with other spectroscopic follow-up resources (TOM or TNS) and surveys (ePESSTO+ among others).

Technical aspects: big data era

- Working efficiently at scale: large stream, data persistence, resiliency, metrics, and auto-remediation.
- Multi-modals analytics capability (streaming & batch) to deal with the complexity of the data, and rarity of interesting events.
- Multi-tenant infrastructure for real-time and post-processing, allowing for as much use cases and users as possible.

3.5M
Observations for 10 years

10K
Alerts per observation

3PB
Alert volume after 10 years

Env. Technical Implementation

- Continuous R&D at CNRS/IN2P3 to explore the big data ecosystem.
  - Main focus on Apache Spark, Kafka, Livy... Cloud-oriented.

- Broker project emerged: Fink.
  - 2 steps architecture: archive vs process.

- Deployed prototypes (alert system & broker) in the VirtualData OpenStack-based Cloud (UPsud). Ongoing discussions with CC.

- Fink tests are promising: up to 7x LSST rate for ingesting data with moderate resources (~50 cores).

- Rich ecosystem in construction: AstroLabNet (analysis frontend), Data monitor (web service), ...

How the community will access the broker outputs?

Envisioned services at CC-IN2P3

Delivering unique set of LSST data products at CC-IN2P3:

- LSST software stack,
- LSST data release,
- *live instance of one LSST broker.*

https://fink-broker.readthedocs.io

Connections

Potential to coordinate efforts with other research institutions and industry

- Include other French institutes (e.g. INSU), with potential broader science cases such as galactic or solar system science
  - Already IAP and Université de Genève.
- Contact computer science researchers who might develop algorithms specially designed for astronomy (e.g. INRIA, PAISS)
- Contact industry partners who can potentially provide support in terms of manpower and hardware
- Expand to European and other international users (e.g., huge TVS community in Italy) - potential to also search for European funding

Next steps

- *15 May 2019: deadline for Letter of Intent for Community Brokers*

- **19-21 June 2019: Workshop in Seattle, USA**
  - May to August 2019: demonstrate the feasibility on a prototype at CC-IN2P3, and develop collaborations.

- August 2019: A few LoI writers will be invited to deliver full proposals

- Summer 2020: Deliver of full proposals

- 2022: Start of LSST operations

Backup

What can we do if we have a broker?

Test, control, select

- Accurate simulations to test selection effects (no intermediaries)
- Rapid response to new situations
- Possibility to acquire target of opportunity spectroscopic follow-up time for very interesting objects

What to avoid when selecting targets

From Rachel Street in the DESC Broker Workshop, Feb 2019

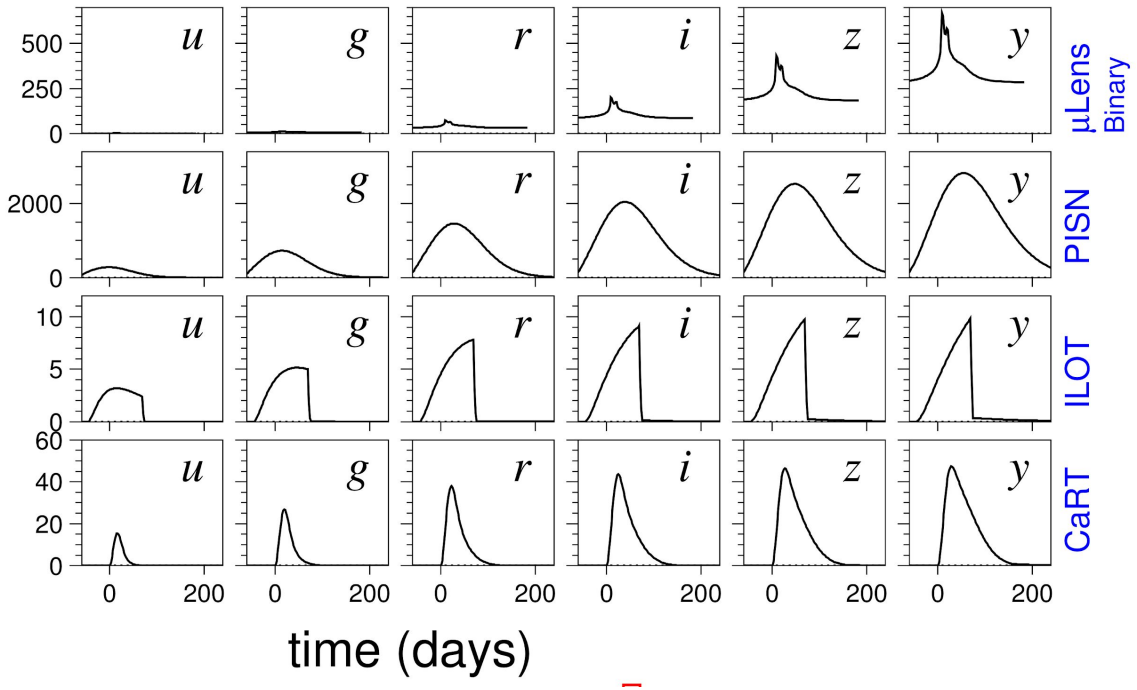
- Understand your selection biases (do not let humans interfere!)
- Avoid Gold-Rush Syndrome

  Everybody is interested in the same objects so facilities are flooded with request for the same objects while interesting others are ignored

Anomaly Detection

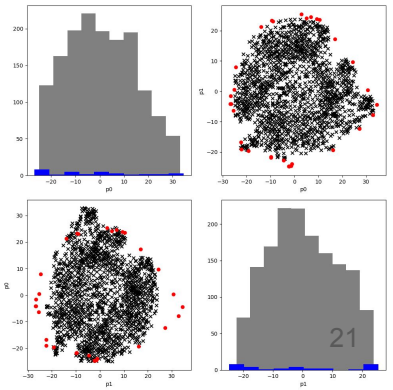
Example of rare or still not detected transient models in the Photometric LSST Astronomical Time-series Classification Challenge (PLAsTiCC)

Anomaly detection not attempted by the top ranked challenge participants



Kessler et al. (incl. Ishida), 2019, PASP - submitted

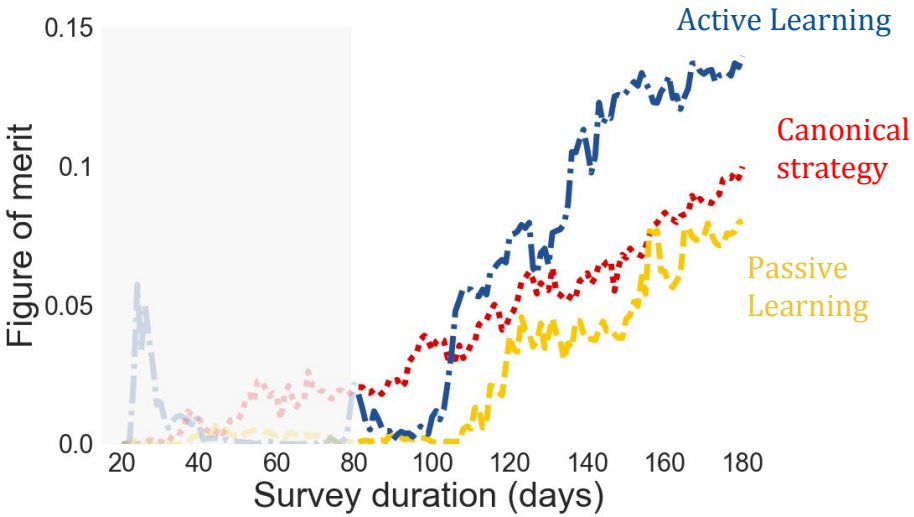
- Potential for discovery of new objects
- Active Anomaly Detection (AAD) can be adapted to optimize the probability to present unusual objects to the expert
- Prototype already working in real data from the Open Supernova Catalog



Ishida et al., 2019 - in prep

Adaptive Machine Learning

for improving classification AND anomaly detection



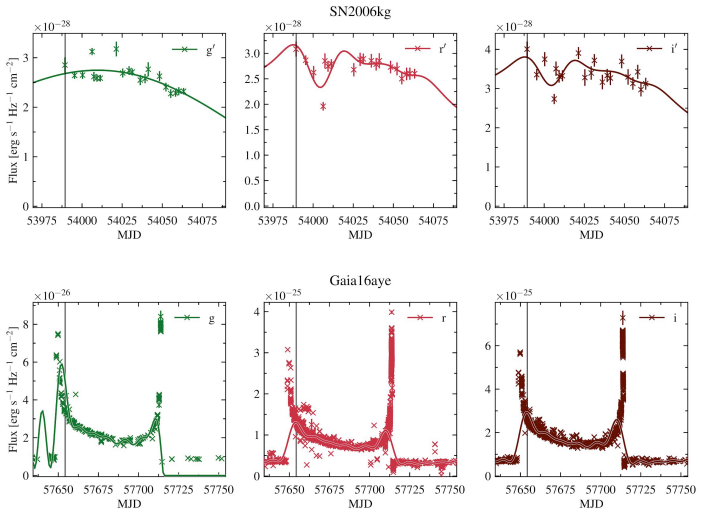
Comparison of supernovae photometric classification results for different spectroscopic follow up strategies.

Results from simulated SNPCC data (Kessler et al., 2010).

Ishida et al., 2019, MNRAS, 483 (1), 2–18

Anomaly detection algorithm automatically identified ~1.5% unusual objects in the OSC

Including 1 AGN <u>miss-classified as SNe</u> (SN2006kg) and the first <u>binary microlensing</u> event discovered in the galactic plane (Gaia16aye)



Ishida et al., 2019 - in prep

Fink: discussion

How can Fink better suit your needs?

- What kind of data products you would like to access from Fink?

- User interface capabilities:
  - Ideas on how to submit your own filter