



*Auvergne, Développement d'Applications et Calcul en Environnement Scientifique
AuDACES, Clermont Ferrand - France, 9 June 2022*



Emille Ishida, Julien Peloton and Anais Möller
on behalf of the Fink Team

The Rubin Observatory Legacy Survey of Space and Time (aka LSST)

In a nutshell:

- telescope: 6.7-m equivalent
- world's largest CCD camera: 3.2 Gpixels

In numbers:

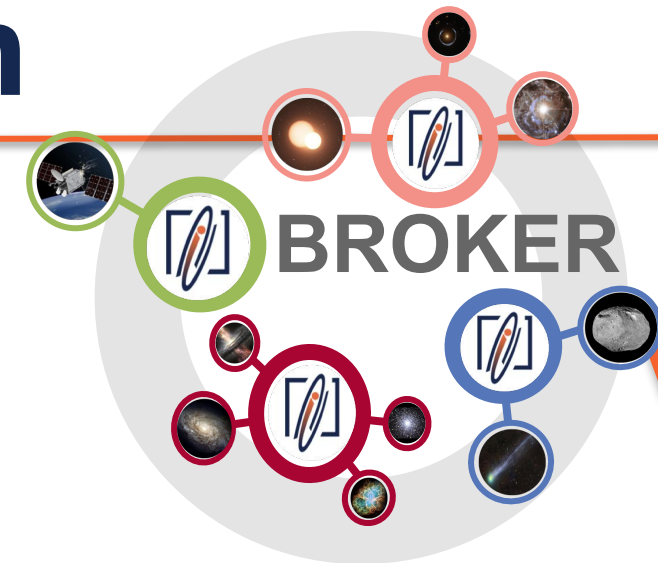
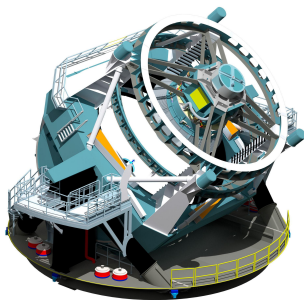
- 10-year survey, starting 2024+
- 1,000 images/night = 15TB/night
- 10 million transient candidates per night
 - Publicly available...
 - ... but huge!



Data path



*every ~30 seconds down to
mag ~24*



*Machine learning
Catalog association
Streams join*

*10 million alerts
per night...*



We would like the interesting ones .3

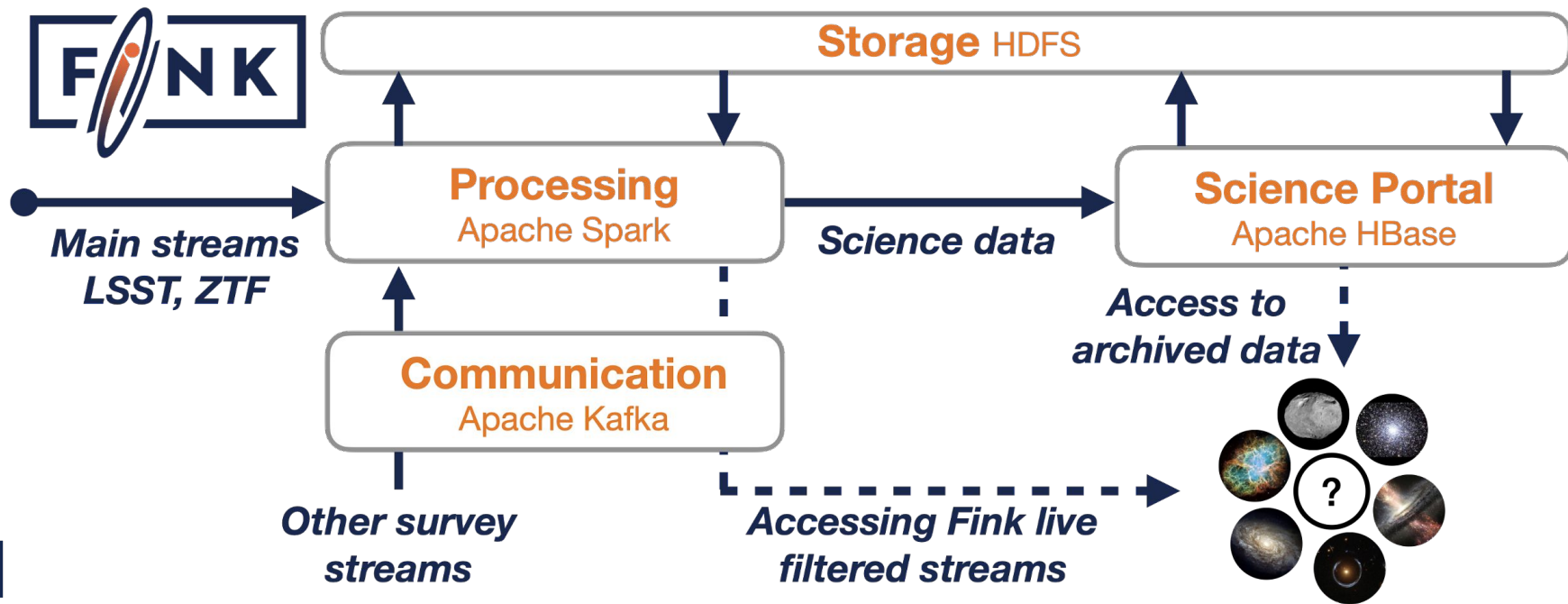
Rubin broker landscape



Fink design

- ✓ Deployed in the cloud (VirtualData, CC-IN2P3*)
- ✓ Collecting alert data from ZTF
- ✓ Benchmarked for LSST data volumes

- ✓ Survey cross-match
- ✓ Public catalogue cross-match
- ✓ Classification (ML, BNNs, Adaptive Learning)

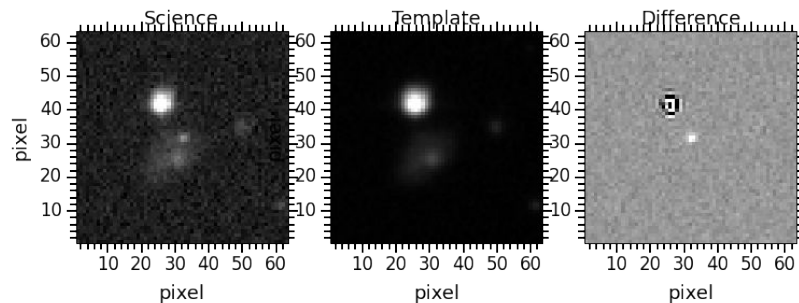
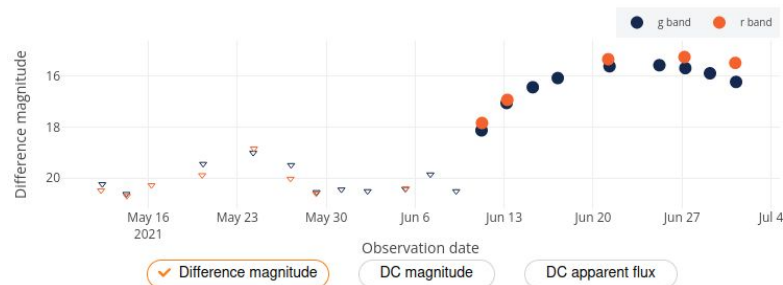


(What is an alert?)

Alerts based on Difference Image Analysis

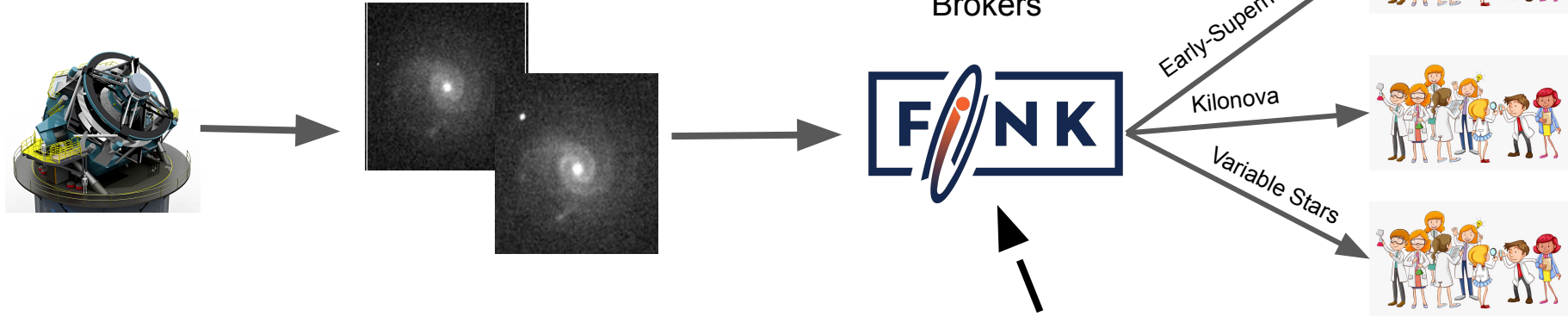
Each alert contains

- Information about the new detection (magnitude, position, ...)
- Neighbours information (xmatches, etc)
- Historical information if the object has been seen previously
- Small images around the detection (60x60 pixels)

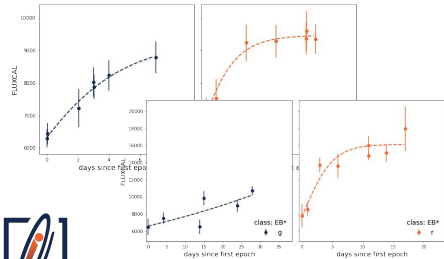


How to classify alerts?

Broker world



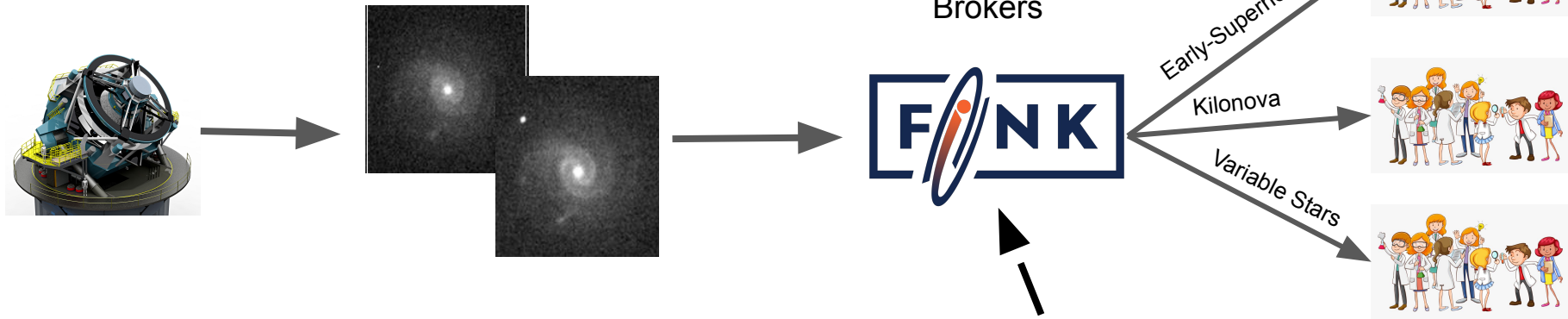
Domain specialist world (this is you)



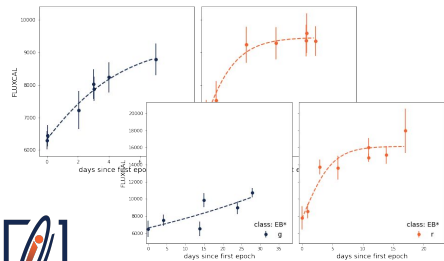
Filter
Catalog xmatch
Taylor science module

How to classify alerts?

Broker world



Domain specialist world (this is you)



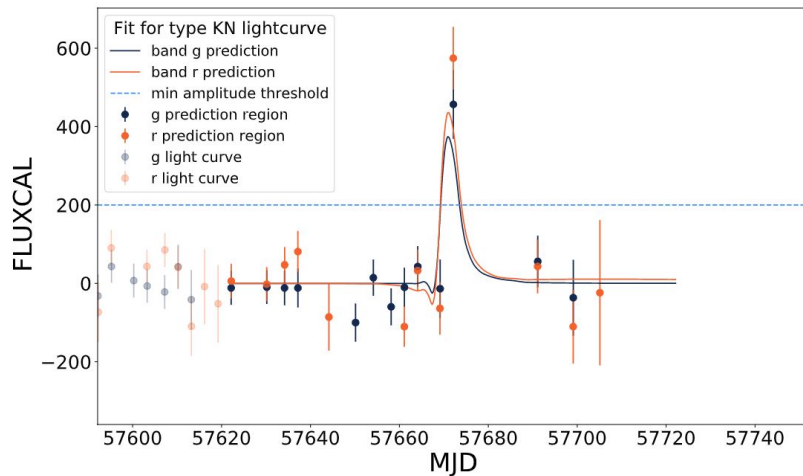
Taylorred science module

$f(\text{alerts}; ++)$ \Rightarrow *class scores*
Boolean

Case study: Kilonova

Problem 1: there are no labels, only 1 confirmed detection

Problem 2: we need to find it fast



(c) KN event, at $z \approx .04$

Data set:

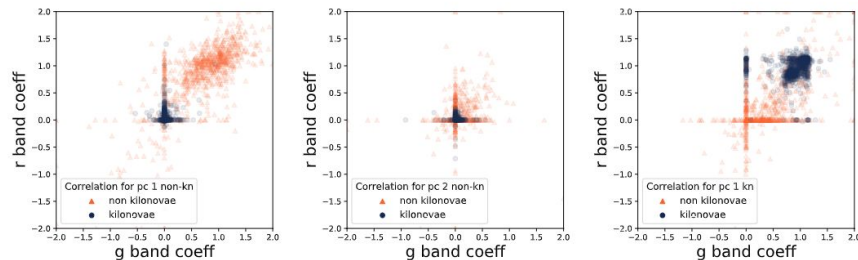
Simulated ZTF light curves

Feature extraction:

Principal components from perfect sims

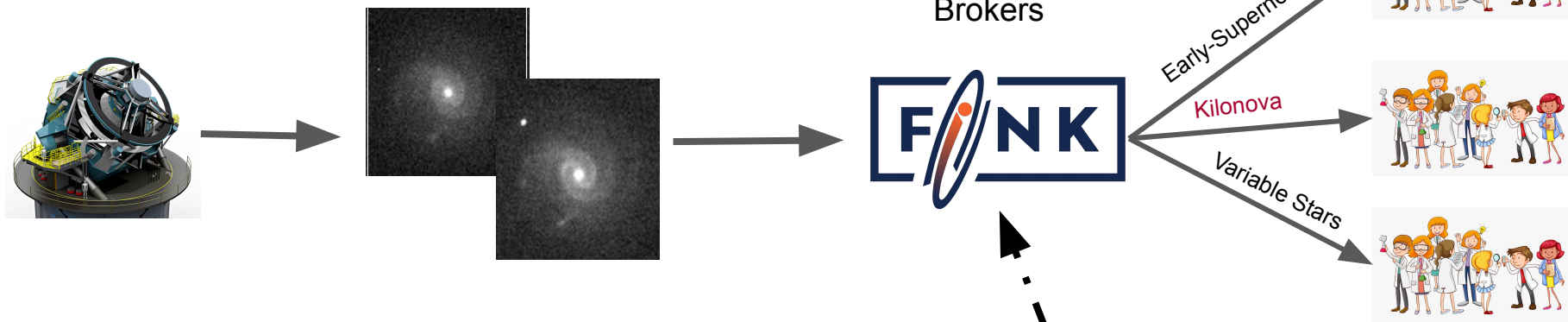
Classifier:

Random Forest

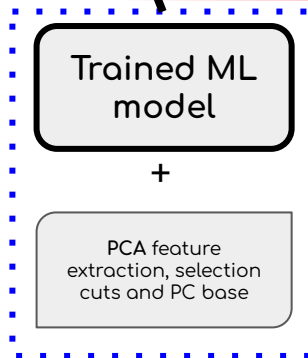
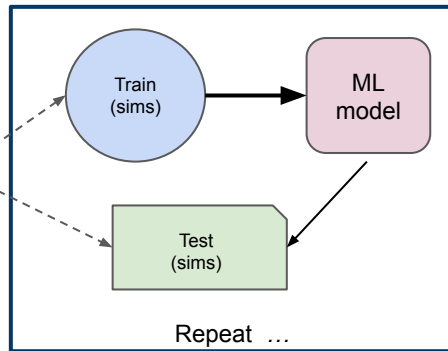
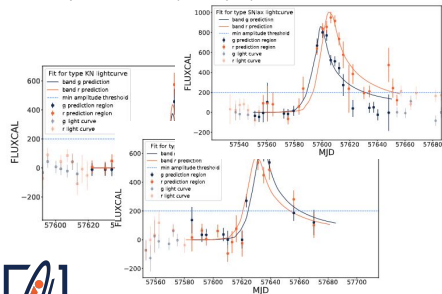


Case study: Kilonova

Broker world



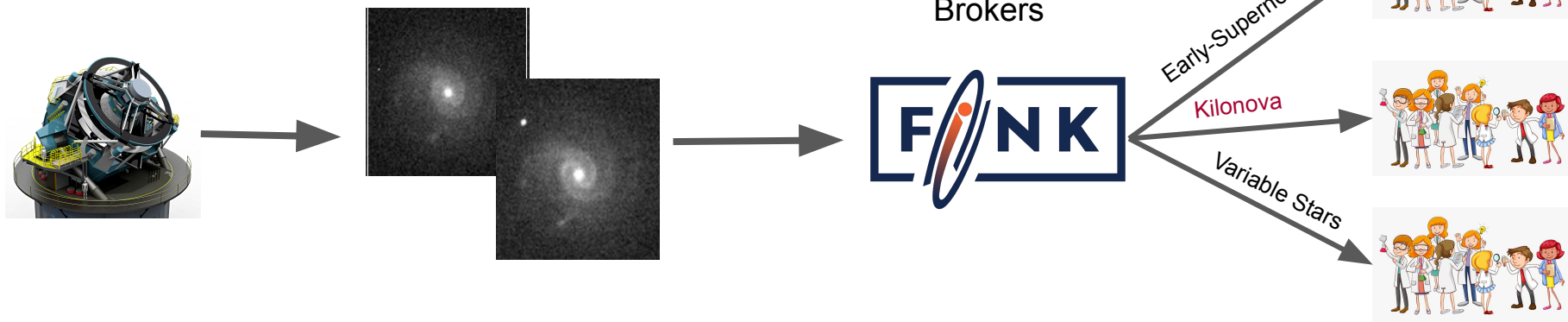
Domain specialist world (this is you)



*Extra imaging from
professional and
amateur
astronomers*

Case study: Kilonova

Broker world



GRANDMA Observations of ZTF/Fink Transients during Summer 2021

Aivazyan et al., 2021, arxiv:astro-ph/2202.09766

- 35 million candidate alerts
- 100 surviving selection cuts
- 6 followed-up by GRANDMA

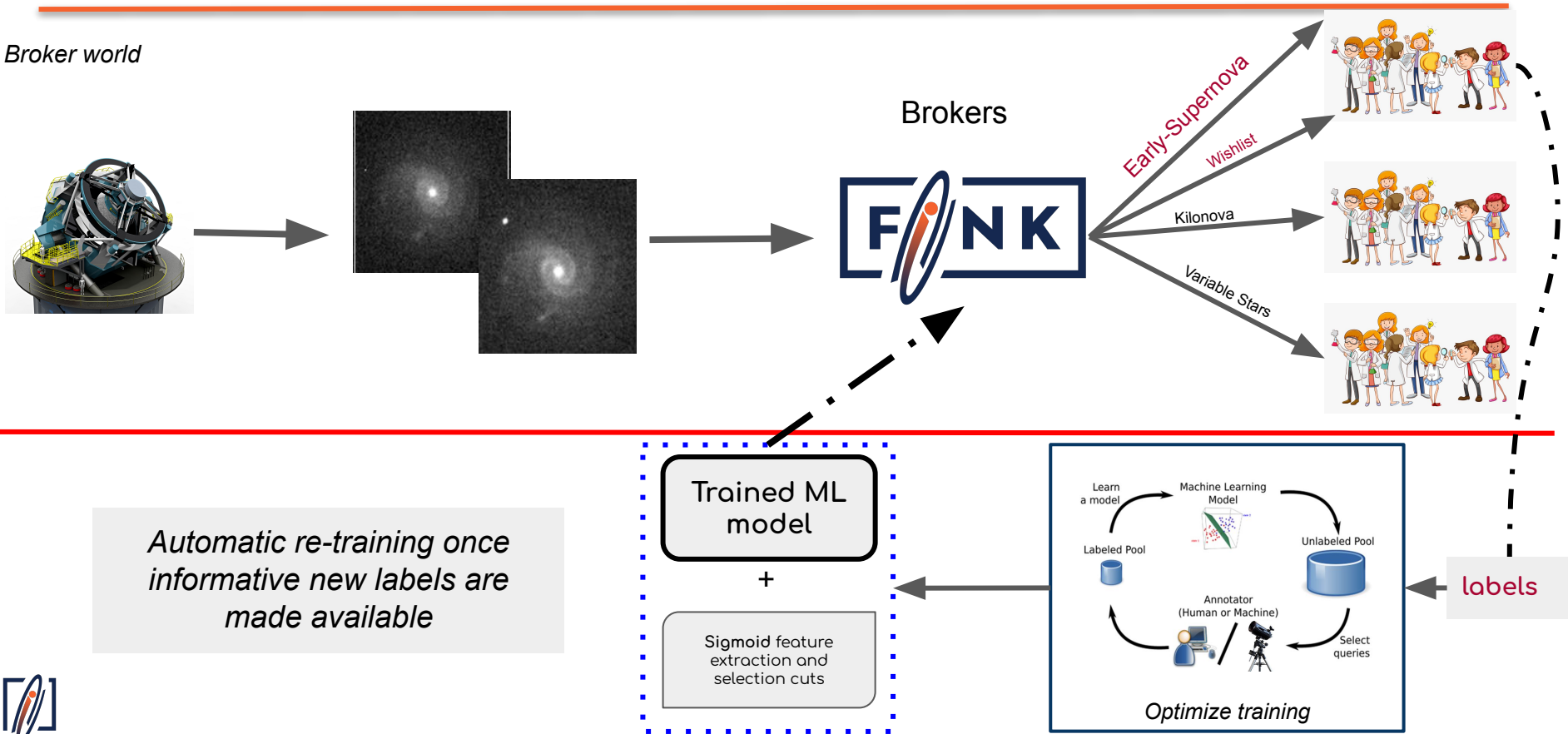
*Extra imaging from
professional and
amateur
astronomers*

Previous talk!

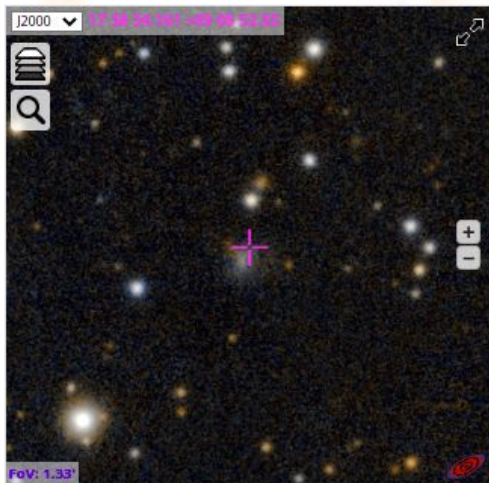


For LSST ...

Broker world



ZTF22aaihxxzh



Inspect alert data

Get object data

Summary

Supernovae

Variable stars

Microlensing

Solar System

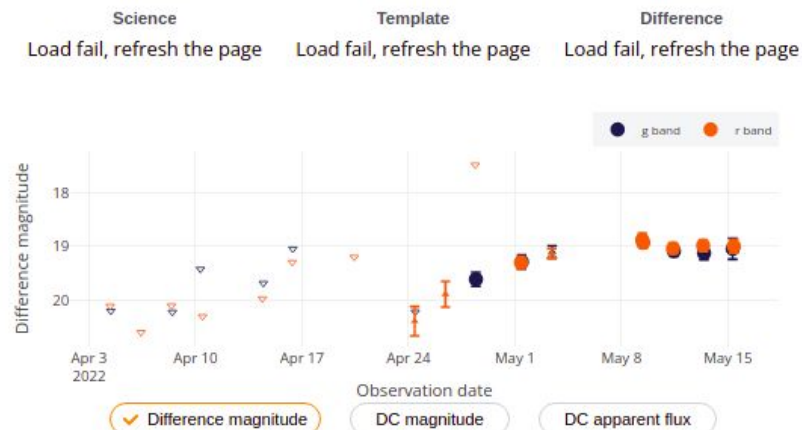
Tracklets

GRB

Individual alert classification

Early SN Ia candidate: 45%

SN candidate: 55%



Information

What am I seeing?

ObjectID: ZTF22aaihxxzh

Last class: Early SN Ia candidate

General properties

Date: 2022-05-15 09:36:47.002

RA: 264.6423366 deg

Dec: 9.1645982 deg

Variability (DC magnitude)

Rate g (last): -0.03 mag/day

Rate r (last): 0.00 mag/day

Neighbourhood

Constellation: Ophiucus

SIMBAD: Unknown

MPC: null

Distance (PS1): 2.91 arcsec

Distance (Gaia): 8.36 arcsec

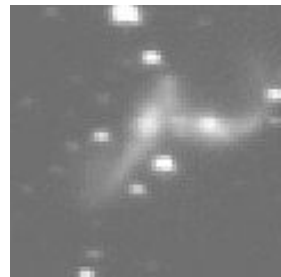
Distance (ZTF): 2.73 arcsec



Accessing Fink data

Two entry points for users:

- Live streams (Kafka streams)
 - Personalisable filters to select objects/parameters of interest
 - Data received “live” (+processing delays)
 - <https://github.com/astrolabsoftware/fink-client>
- Science Portal & REST API
 - All data will remain accessible for the full survey duration
 - <https://fink-portal.org>
- TOM module
 - https://github.com/TOMToolkit/tom_fink
- Statistics information:
 - <https://fink-portal.org/stats>



Take home message

- Preparing for LSST means be prepared to define what is interesting
- Automatization of recommendation systems can enable improved classification and discovery
- Fink was specifically designed to enable incorporation of complex queries (domain knowledge)



The best of Fink ...



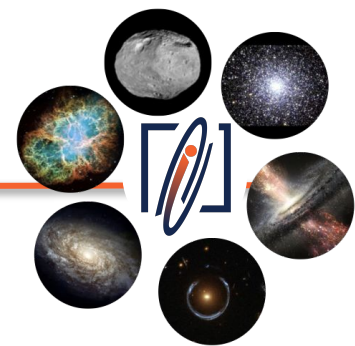
... are the people
behind it!

1st Fink collaboration meeting
LAPP, 19 - 20 May 2022

<https://indico.in2p3.fr/event/26707/>



You are welcome to join!



- Fink is a broker designed specifically for LSST
- Enabling science by applying state-of-the-art technology.
- Technology Readiness Level (TRL) 6/9.
- Currently digesting ZTF stream

First science modules deployed and testing capabilities beyond expectations: SNe, GRB, KNe, microlensing, ...

New proposals for science modules are welcome! <https://fink-broker.org/joining.html>

More info:

- Fink white paper, [arXiv: astro-ph/2009.10185](https://arxiv.org/abs/astro-ph/2009.10185)
- Website & Science Portal: <https://fink-broker.org>
- API Tutorials: <https://github.com/broker-workshop/tutorials/tree/main/fink>





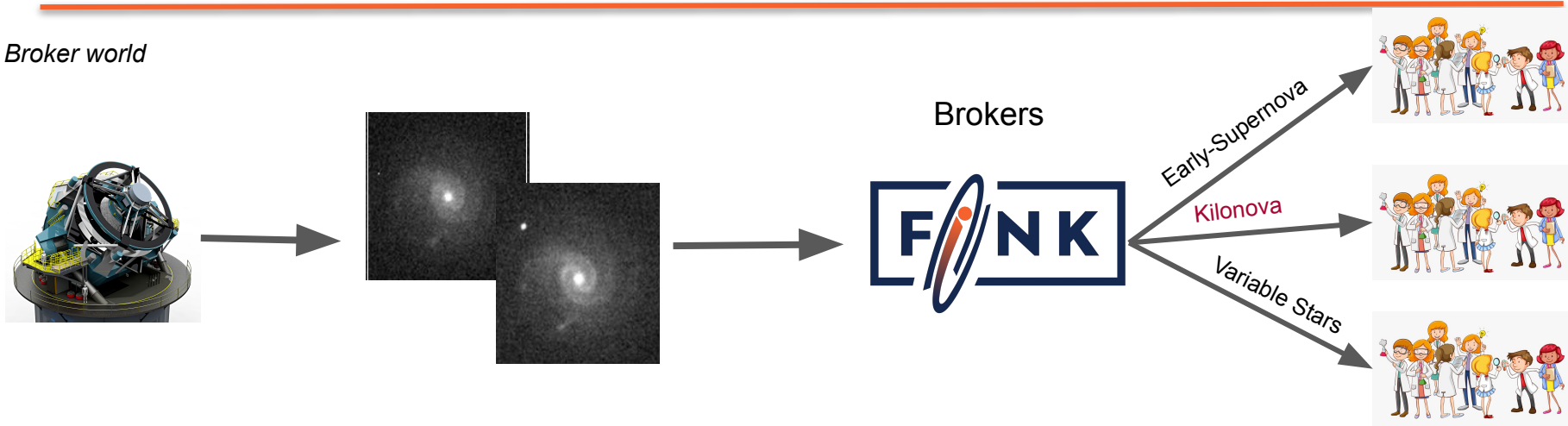






Case study: Kilonova

Broker world



GRANDMA Observations of ZTF/Fink Transients during Summer 2021

Aivazyan et al., 2021, arxiv:astro-ph/2202.09766

- 35 million candidate alerts
- 100 surviving selection cuts
- 6 followed-up by GRANDMA

*Extra imaging from
professional and
amateur
astronomers*



To remember

- Fertile ground for machine learning applications and **interdisciplinarity**
- Designed to accommodate **deep and adaptive** learning methods
- **Cross-match** with LSST images for users with data rights

*First science modules deployed and testing capabilities beyond expectations:
SNe, GRB, KNe, microlensing, ...*

More info:

- Fink white paper, [arXiv: astro-ph/2009.10185](https://arxiv.org/abs/astro-ph/2009.10185)
- Website & Science Portal: <https://fink-broker.org>
- API Tutorials: <https://github.com/broker-workshop/tutorials/tree/main/fink>

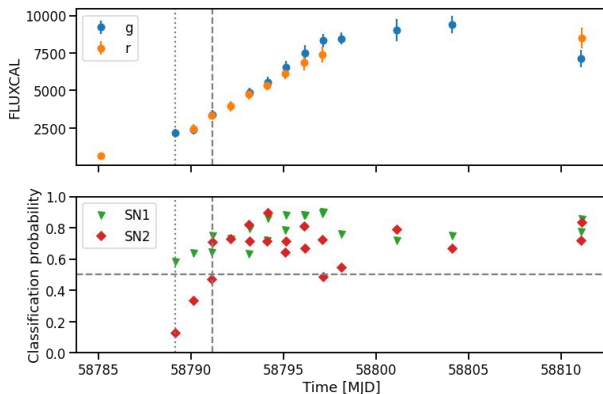


Example: SN classification

High-accuracies using *Deep Learning*

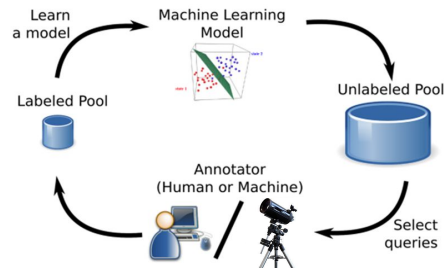
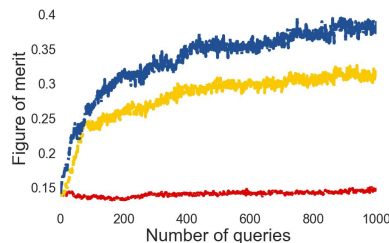


Moller and Boissiere, 2020, MNRAS



- All science modules are built by the community
- Outputs from all science modules will be publicly available
- Hosted at CC-IN2P3: cross-match with LSST image data available for users with data rights
- Distributed Machine Learning

Increasingly more accurate classifications and anomaly scores using *Adaptive Learning*



Ishida et al., 2019, MNRAS

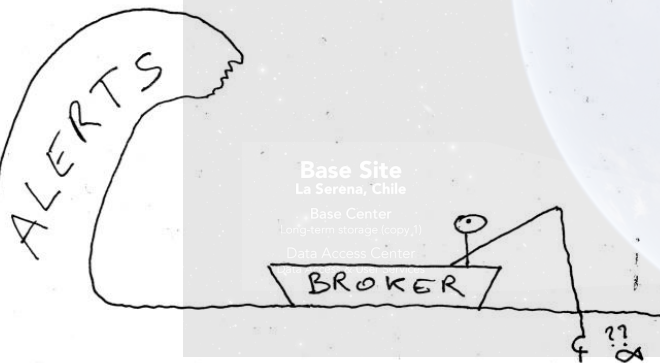


Alert data challenge

Forecasted: **10 million alerts per night...**

- Current serialisation implies ~100KB/alert, **1TB/night**, 3PB in 2030.

Fink is made for LSST



LSST
Large Synoptic Survey Telescope

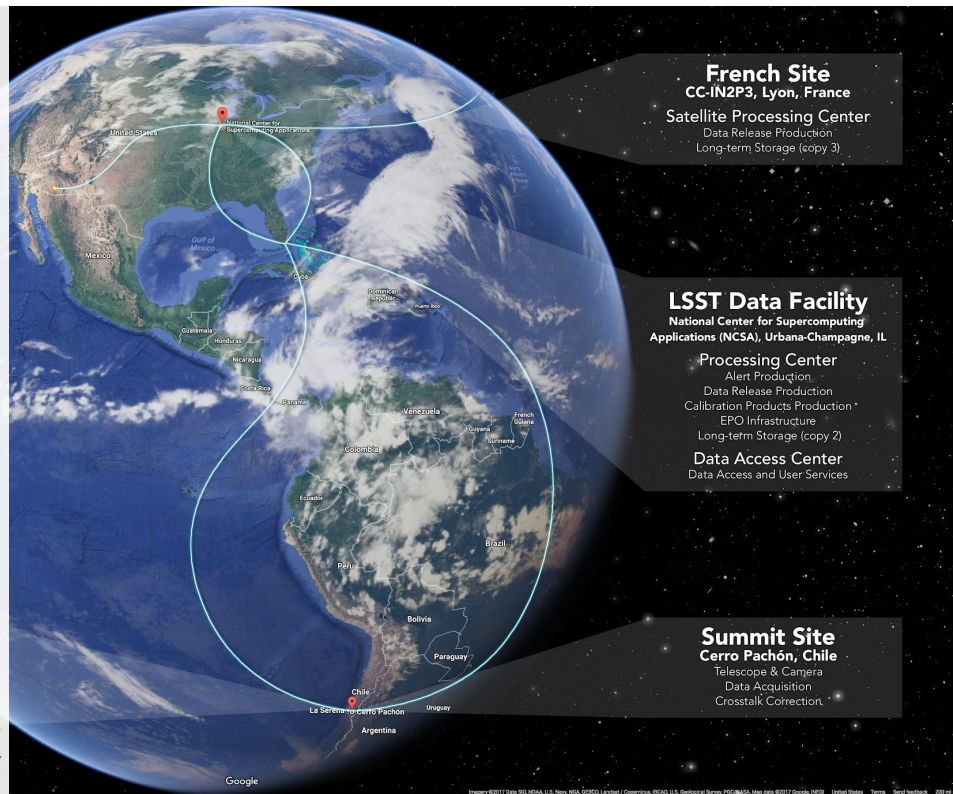
LSST Operations:
Science & Data Flow

HQ Site
Tucson, AZ

Science Operations
Observatory Management
Education & Public Outreach

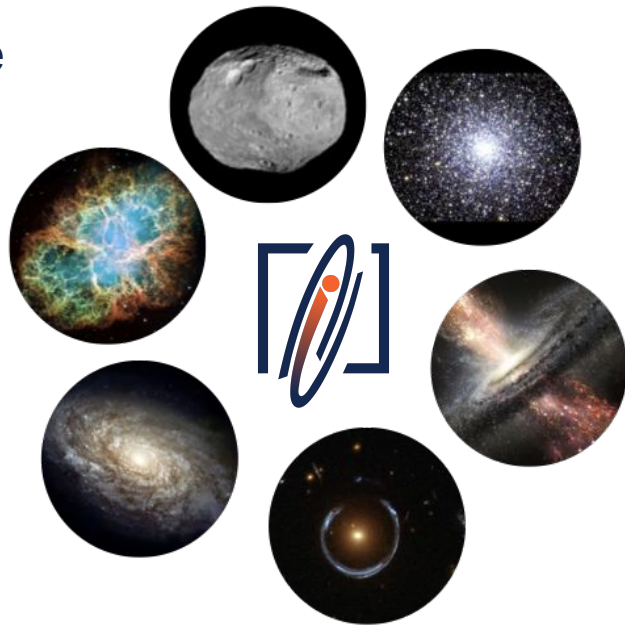
Base Site
La Serena, Chile

Base Center
Long-term storage (copy 1)
Data Access Center



Fink scientific objectives

- Fink is a community-driven effort, open to anyone
- Current fields of expertise
 - Supernovae
 - Kilonovae
 - Microlensing
 - Multi-messenger astronomy
 - GRB, X, neutrino, GW...
 - Anomaly detection
- We are open to new contributions!
 - Recent: Solar System Objects
 - ... *<your project here!>*



<https://fink-broker.org/joining.html>

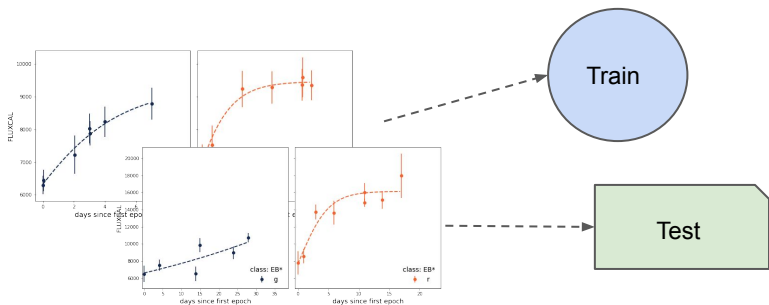


Case study: Early SN Ia classification

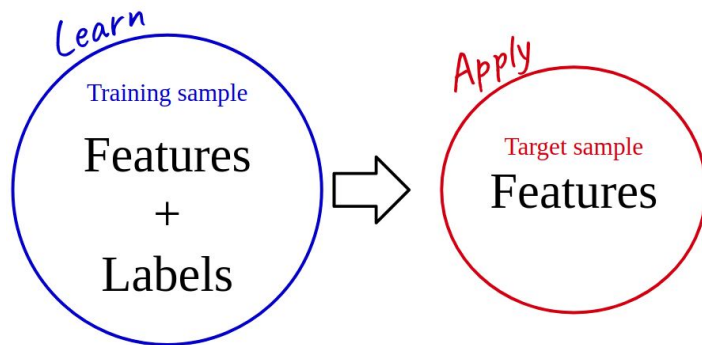
Problem 1: labels are expensive, resources are limited

Problem 2: training (spectroscopically classified light curves) is not representative from test (purely photometric light curves)

Goal: optimize classification results with small training



Ideal data situation

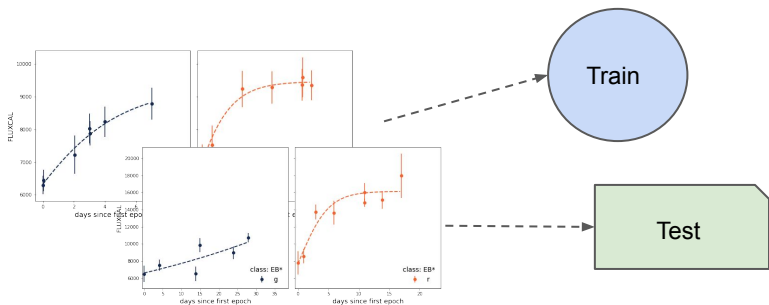


Case study: Early SN Ia classification

Problem 1: labels are expensive, resources are limited

Problem 2: training (spectroscopically classified light curves) is not representative from test (purely photometric light curves)

Goal: optimize classification results with small training



Real data situation

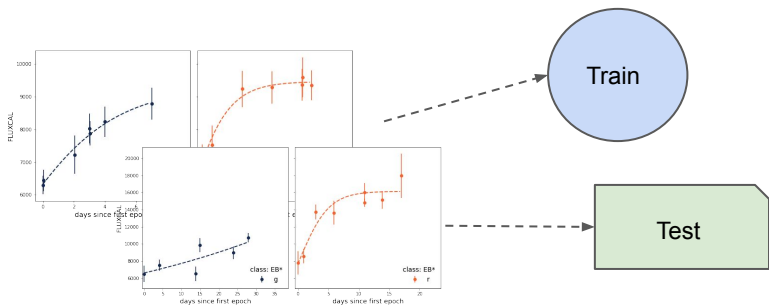


Case study: Early SN Ia classification

Problem 1: labels are expensive, resources are limited

Problem 2: training (spectroscopically classified light curves) is not representative from test (purely photometric light curves)

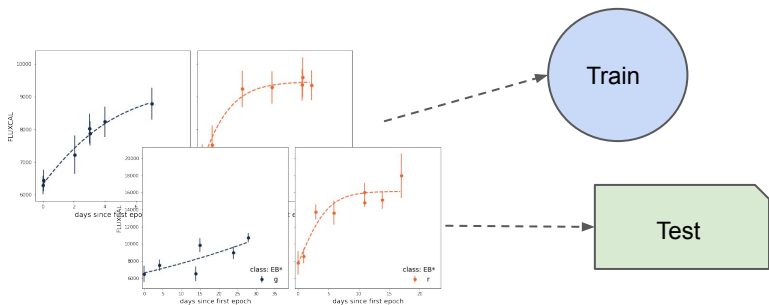
Goal: optimize classification results with small training



Case study: Early SN Ia classification

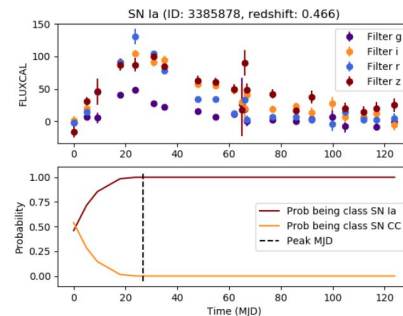
Problem 1: labels are expensive, resources are limited

Problem 2: training (spectroscopically classified light curves) is not representative from test (purely photometric light curves)



Goal: optimize classification results with small training

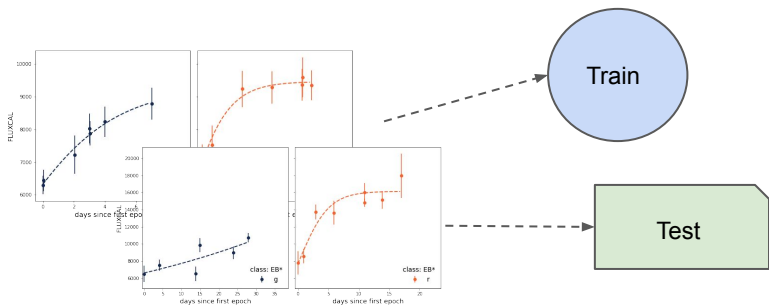
Strategy 1: extract information from sims using BNN



Case study: Early SN Ia classification

Problem 1: labels are expensive, resources are limited

Problem 2: training (spectroscopically classified light curves) is not representative from test (purely photometric light curves)



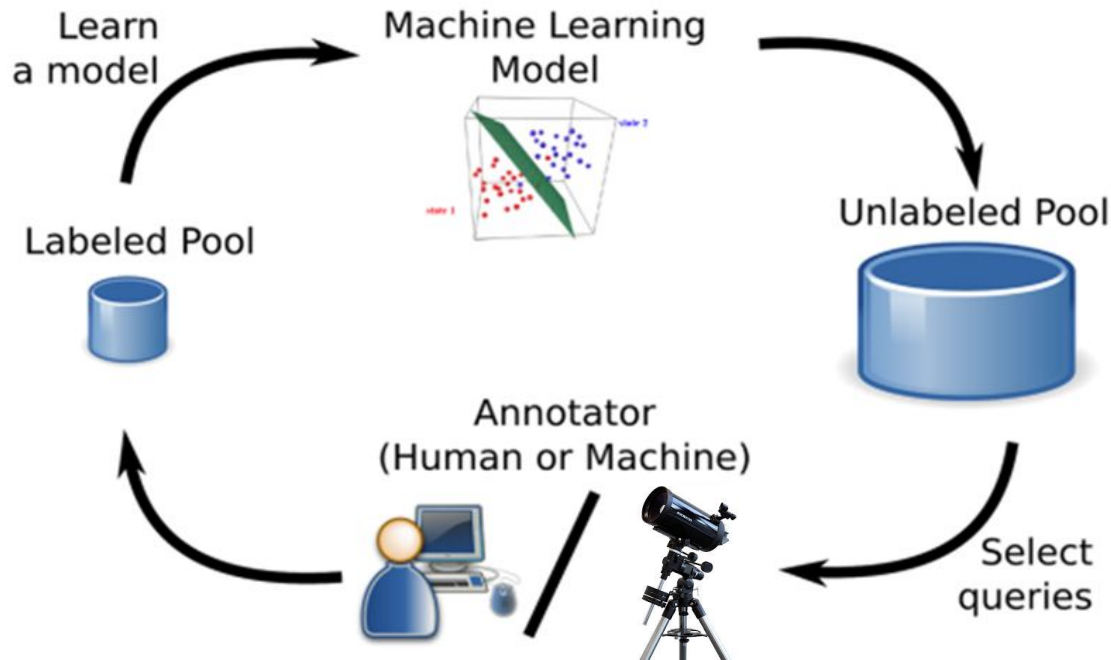
Goal: optimize classification results with small training

Strategy: optimize the construction of training samples

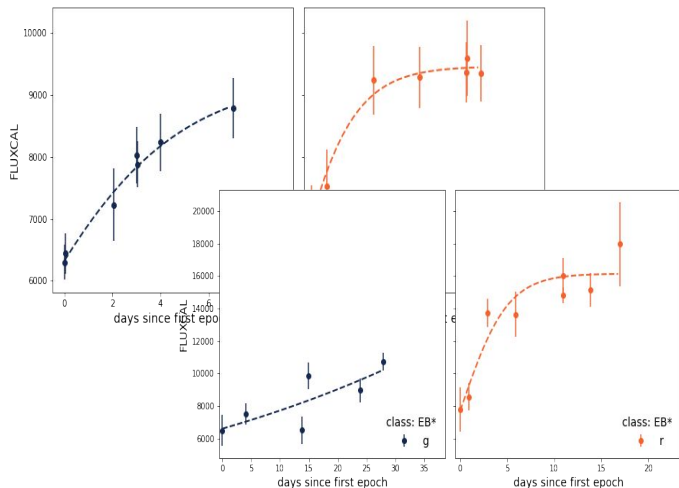


Case study: Early SN Ia classification

*Active Learning
Optimal experiment design*



Case study: Early SN Ia classification



Data set:

52 673 alerts with labels from TNS or simbad, of which
3 033 were SN Ia (~7%)

Initial training:

10 alerts (5 SNIa, 5 others)

Feature extraction:

Sigmoid fit, 3 rising points

Goodness of fit, SNR, number of points after cuts

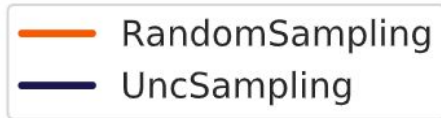
After learning loop:

310 in training

52 363 in test

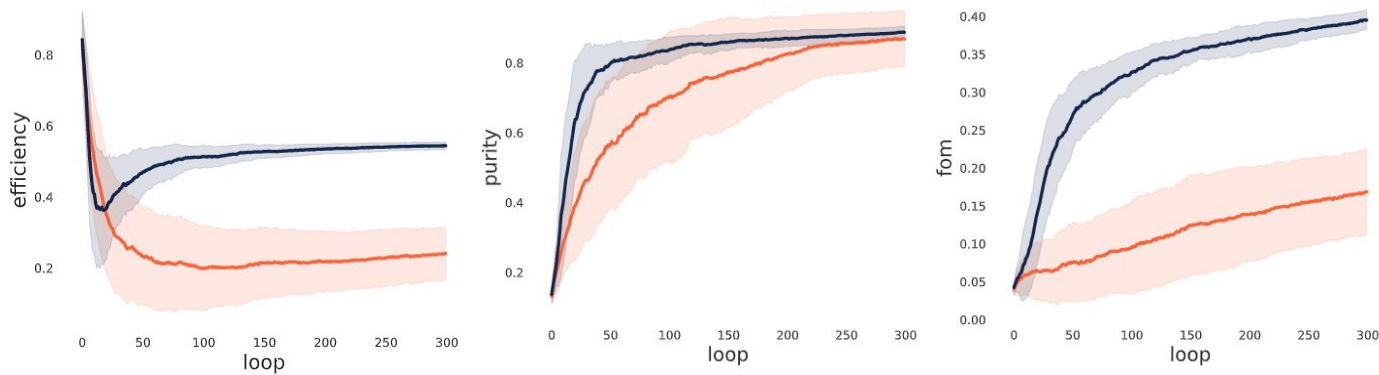


Case study: Early SN Ia classification



Results after 300 loops:

Training: 310 alerts
Testing: > 52 000 alerts



Choose training sample which lead to better results
and train a Random Forest classifier ...

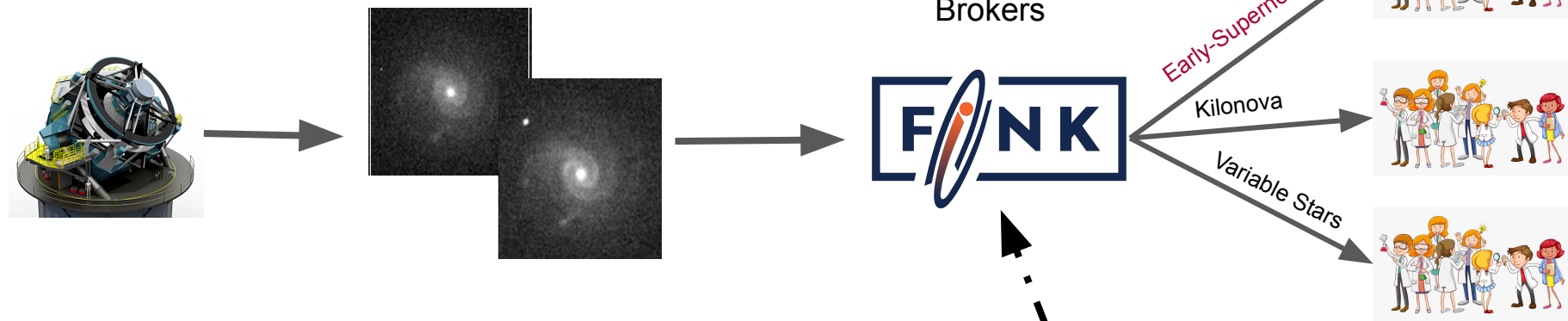


Trained ML
model

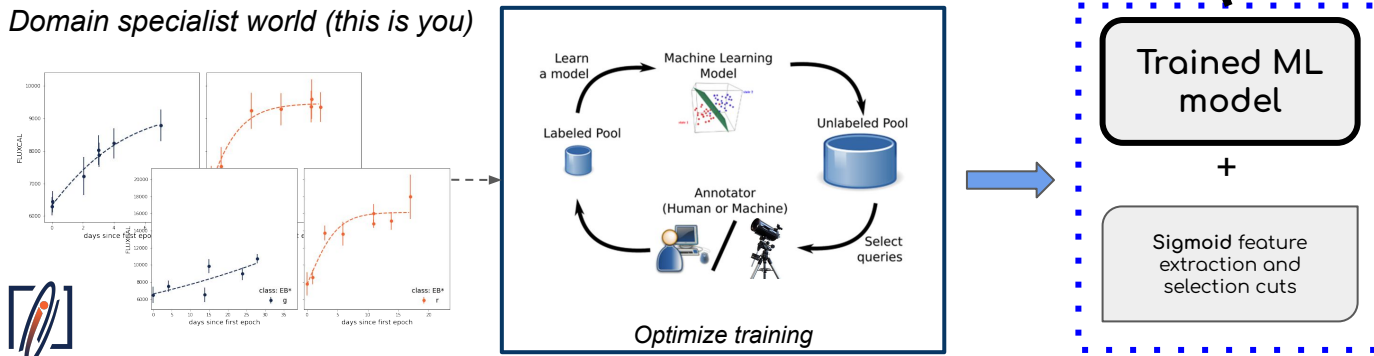


Case study: Early SN Ia classification

Broker world

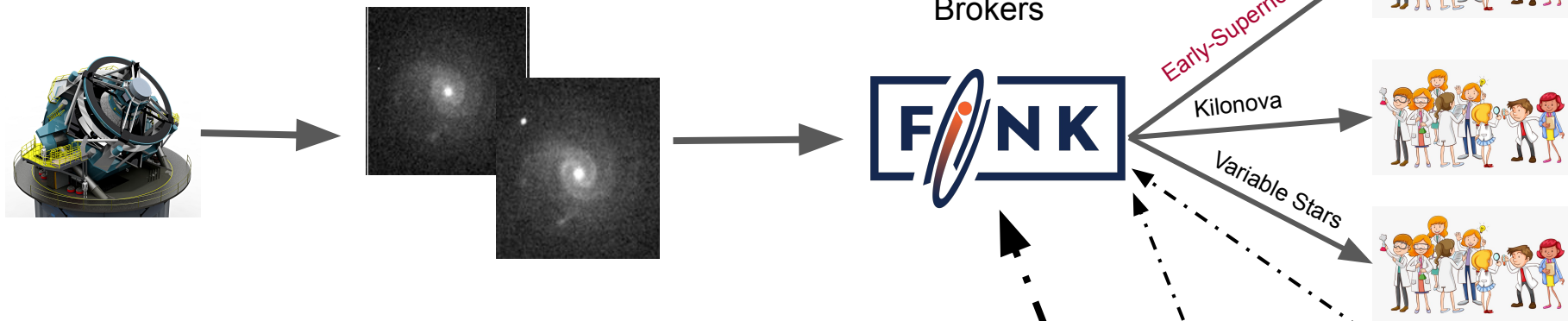


Domain specialist world (this is you)

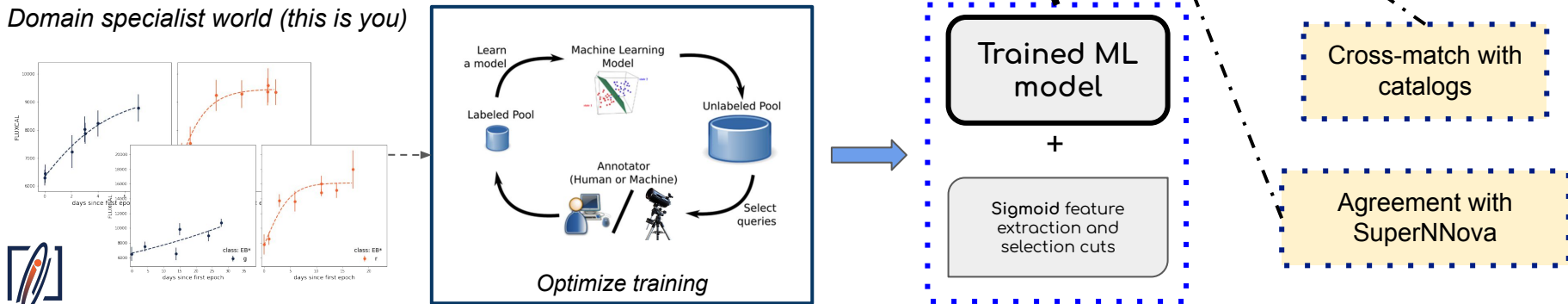


Case study: Early SN Ia classification

Broker world



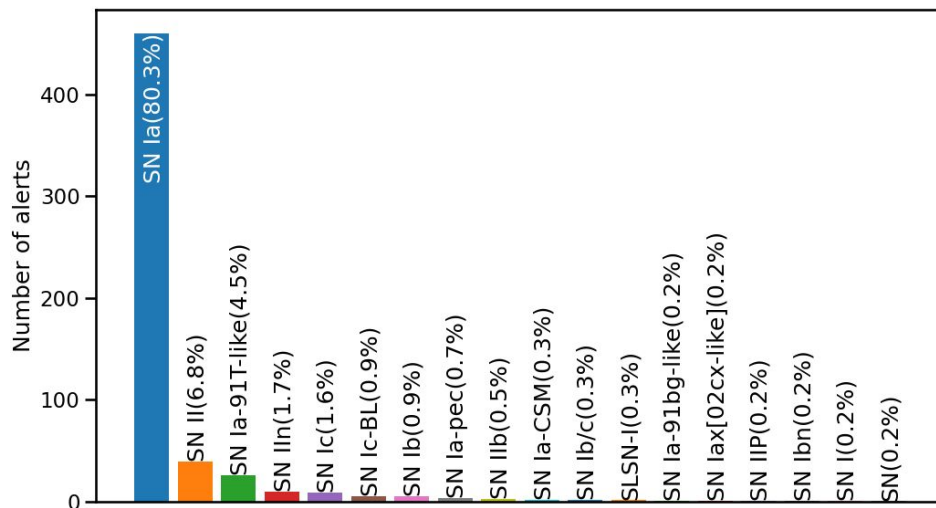
Domain specialist world (this is you)



Case study: SN classification

Fink Early SN Ia candidates reported to TNS from November/2020 - March/2022:

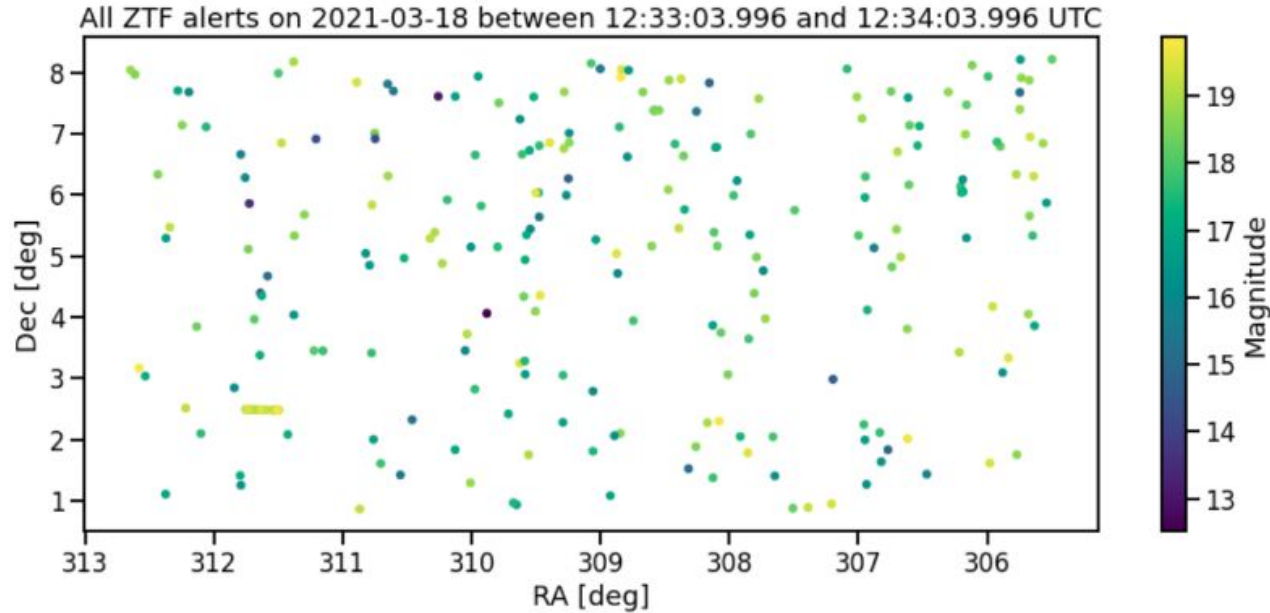
- 4 661 Early SN Ia candidates
- 573 spectroscopically classified
- Contaminants are mostly other SNe
 - 1 LBV



Plot by Julien Peloton (CNRS/IJCLab)



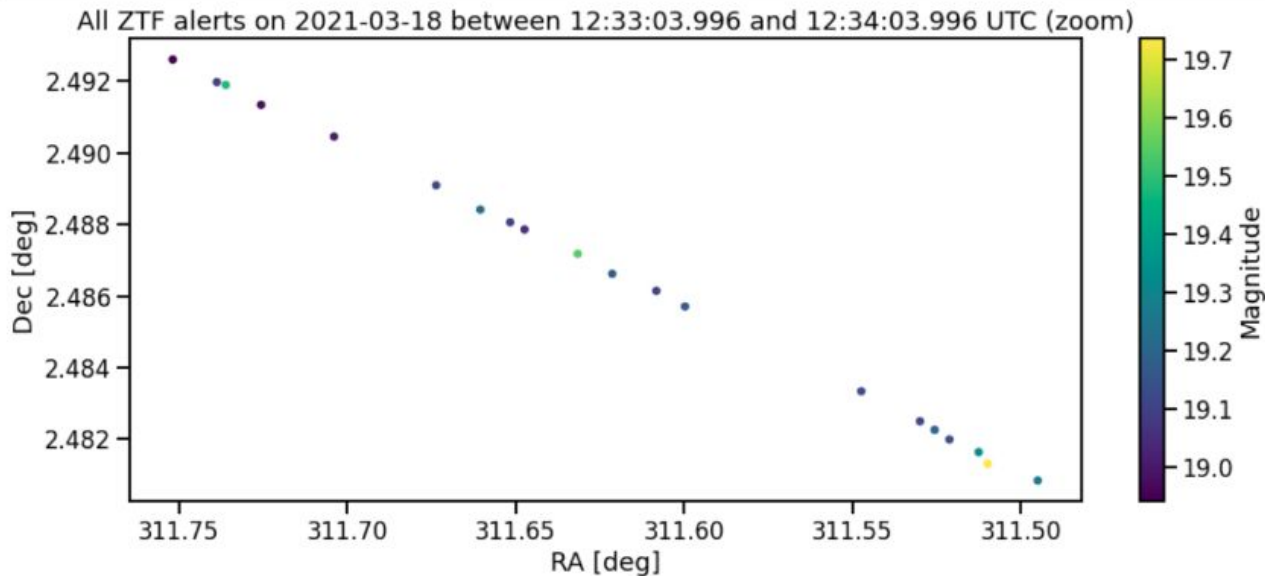
Curiosities: the mystery



Alerts in 1 exposure



Curiosities: the mystery

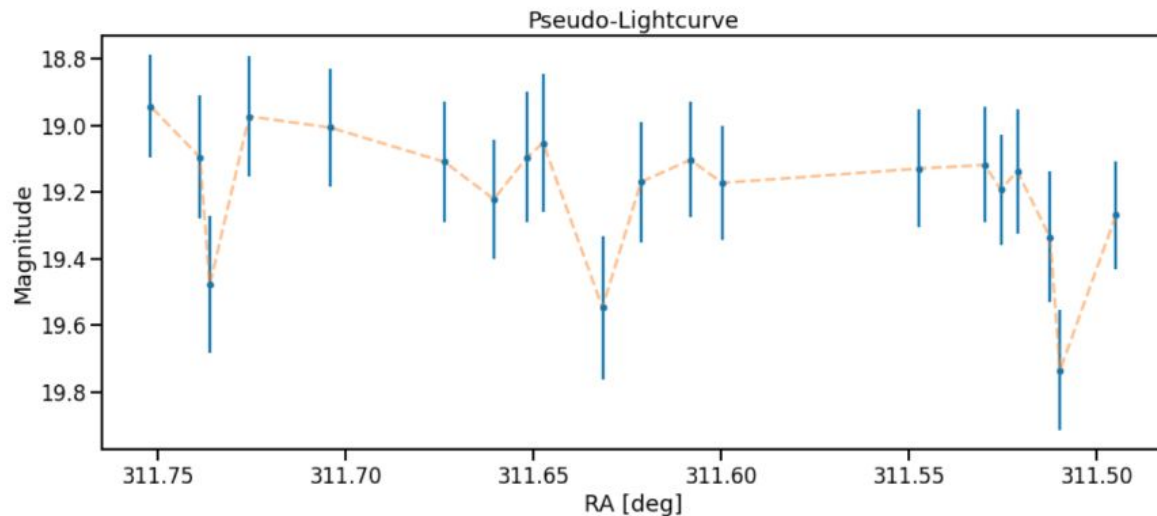


Alerts in 1 exposure

Very fast brightness variation



Curiosities: the mystery



Alerts in 1 exposure

Very fast brightness variation

Possibilities:

- Spy satellite?
- Space debris?
- Alien spaceship ?



Current project highlights

- ❑ **Long** transients (Blodin, Pruzhinskaya)
- ❑ AL for **early SN Ia** discovery (Leoni)
- ❑ **Anomaly** detection (Pruzhinskaya, Kornilov, Russeil, Beschastnov)
- ❑ Detection and classification of **satellite glints & debris** (Karpov)
- ❑ Search for **GRB** afterglows (+orphans) (Bregeon, Turpin, Ducoin, Le Montagner)
- ❑ Search for **Kilonova** (Biswas, GRANDMA collaboration)
- ❑ New techniques to discover **SSO** objects (Le Montagner)
- ❑ New database techniques using **graphs** (Hrivnac)

