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

MMA and discovery with Rubin

- State of the art data sets in optical astronomy
- The Rubin broker ecosystem
- The role of Machine Learning
- The human factor
How did we get here?

Big data happened

SDSS
2000 - now
Primary mirror: 2.5 m
120 TB

ZTF
2018 - now
Primary mirror: 1.2 m
1.4 TB/night

From 2025
Primary mirror: 8.4 m
15 TB/night
The Vera Rubin Observatory
Large Survey of Space and Time (LSST)

In a nutshell:
- telescope: 8.4m primary mirror
- 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!
To keep in mind …

1) Very big data

2) Survey mode observation strategy

How to distribute this to people who will do science?
From detection to science

The data path

10 million alerts per night...

We would like the interesting ones ...
Rubin broker landscape
(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

Brokers

TDE

Kilonova

Anomalies

Domain specialist world (this is you)
How to classify alerts?

Broker world

You can access this via de Fink Science portal or the API

All alerts data is public!

https://fink-portal.org/

Domain specialist world (this is you)
How to classify alerts?

Broker world

Domain specialist world (the expert)

Filter
Catalog or stream xmatch
Taylored science module
How to classify alerts?

Broker world

Domain specialist world (the expert)

- Magnitude
- Position, or any column really
- X-match with
  - existing catalogs
  - other streams
  - probability maps
Xmatch with GW streams

O4 is coming – Fink has already some tool to play with GW sky maps

https://fink-portal.org/api → Gravitational Waves → tutorial!

```python
# Query Fink
data = gzip.open(fn, 'rb').read()
r = requests.post(
    'https://fink-portal.org/api/v1/bayestar',
    json={
        'bayestar': str(data),
        'credible_level': credible_level,
        'output-format': 'json'
    }
)
```
How to classify alerts?

Broker world

Domain specialist world (the expert)

Taylored science module

\[ f(\text{alerts}; ++) \Rightarrow \text{class scores Boolean} \]
How to classify alerts?

Broker world

Domain specialist world (the expert)

Trained ML model +

Feature extraction, selection cuts, auxiliary data, etc...

You can also use any other information already available in the broker

Repeat …

Train

ML model

Test

Repeat …
Example: TDE

- Focused on rising TDE examples
- Remove things we already know (xmatch with whatever possible)
- Multi-wavelength feature extraction with Rainbow
- Anomaly detection based ML-model

Work by Erwan Quintin, Miguel Llamas Lanza and Etienne Russell
Receiving your candidates

- Kafka stream -- for on-the-fly notifications
Receiving your candidates

- Bots

**Cross-match-based kilonova bot**

Fink Science Portal: ZTF24aaemydm

**SkyPortal:** ZTF24aaemydm

**Time:**
- 2024-02-12 08:02:33.003 UTC
- Time since first detection: 1.9 hours

**RA/Dec:**
- [hours, deg]: 9 54 28.51 -18 38 10.9
- [deg, deg]: 148.6188113 -18.6363575

**Galactic latitude:**
- [deg]: 27.2874401

**TNS:** link

**Presumed host galaxy:**
- HyperLEDA Name: NGC3052
- 2MASS XSC Name: 09542791-1838202
- Luminosity distance: (57.25 ± nan) Mpc
- RA/Dec: 148.6163330 -18.6389450
- log10(Stellar mass/Ms): 10.81

**Cross-match:**
- Alert-host distance: 3.43 kpc
- Absolute magnitude: -15.54

**Measurement (band r):**
- Apparent magnitude: 18.25 ± 0.07
Receiving your candidates

- Added value + download service

Fink Data Transfer

Data Source
Choose the type of alerts you want to retrieve
- ZTF  □  ELASTICC (v1)  □  ELASTICC (v2.0)  □  ELASTICC (v2.1)

Date Range *
Pick up start and stop dates (included).
- June 1, 2023 – January 31, 2024

Alert class
Select all classes you like! Default is all classes.
- (Fink) Supernova candidates  △  (TNS) SN  △  (SIMBAD) SN

Extra conditions
One condition per line (SQL syntax), ending with semi-colon. See here for fields description and here for examples.
- nalert>5; magpsf < 22

Alert content
Choose the content you want to retrieve
- Lightcurve (~1.4 KB/alert)  □  Cutouts (~41 KB/alert)  □  Full packet (~55 KB/alert)

Submit Job

https://fink-portal.org/download
The beauty of an observational science

“... telescopes that merely achieve their stated science goals have probably failed to capture the most important scientific discoveries available to them.”

Fink Anomaly Detection module

- Started with extragalactic experts, now Fink-wide engagement
- Random forest model
- Quick reaction from the follow-up community:
  - 9.2m SALT (South Africa)
  - 0.6m and 2.5m KGO (Russia)
  - 0.25m FRAM-ORM (Spain)
  - 0.2M FOSC-ES32 (Italy)

**FIRST NIGHT – FIRST ANOMALY**

https://fink-portal.org/ZTF23aaaatw1

- ASASSN-23ac/PNV J06245297+0208207
- Simbad: WD candidate

REPORTED IN TNS AS AT 2023AWT
Fink Anomaly Detection module

Active anomaly discovery in the alert stream for personalized anomaly detection models

Implementation by the SNAD team, via coniferest package
Fink Anomaly Detection module

Active anomaly discovery in the alert stream for personalized anomaly detection models

Implementation by the SNAD team, via coniferest package
Fink Anomaly Detection module

Active anomaly discovery in the alert stream for personalized anomaly detection models

Implementation by the SNAD team, via coniferest package

Fink anomaly bot
Median anomaly score overnight: 0.01.
Personalized ML for big data

Fink Collaboration meeting 2024

Fink hackathon 2022

Fink collaboration meeting 2022

FinkBR 2024

ozFink meeting 2023
Get inspired

#FinkDreamShots

- Build a catalogue of interstellar asteroids
- Find a white hole?
- A classifier for tidal disruption events
- I want to find a live PISN
- I wish astronomers use REAL units!
- A switch between magnitudes and luminosities

From OzFink 2023 - Melbourne, Australia - https://www.ozgrav.org/ozfink-workshop-2023.html
What do you want to see?
Case study: Kilonova

Problem 1: there are no labels, only 1 confirmed detection - with a GW counterpart

Problem 2: we need to find it fast

Data set:
- Simulated ZTF light curves

Feature extraction:
- Principal components from perfect sims

Classifier:
- Random Forest

https://fink-portal.org/ZTF21abgcgyq

Case study: Kilonova

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
Case study: Early SN Ia classification

Active Learning
Optimal experiment design

Plot modified from Chowdhury et al., 2021, SPIE Medical Imaging
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

Leoni et al., arxiv:2111.11438, in press A&A
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

**Case study:** Early SN Ia classification

![Plot by Julien Peloton (CNRS/IJCLab)]
Active Anomaly Detection

Plot modified from Chowdhury et al., 2021, SPIE Medical Imaging
Case study: Satellite tracks

**Problem 1:** they hide in plain sight. labels must evolve

**Problem 2:** they move fast and may confuse difference image analysis

- Module to identify satellite glints
- 11.5% all single-frame events
- 30% of those with real-bogus > 0.8
- 140 per night