

MANTRA: A reference dataset for astronomical transient event recognition.

Part I: lightcurves

Mauricio Neira
Juan Pablo Reyes

Catalina Gómez
Marcela Hernández Hoyos
Jaime E. Forero-Romero

Diego A. Gómez
Pablo Arbeláez

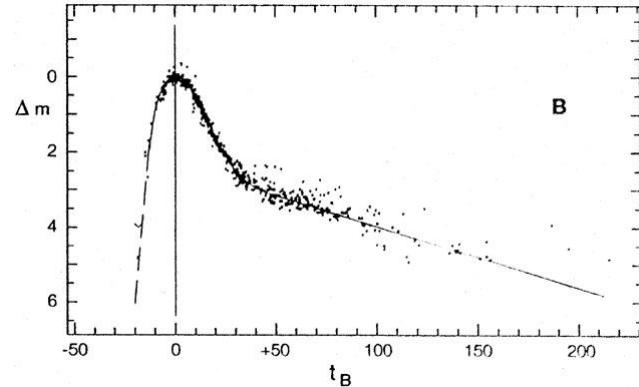
Problem

Classification of Transient Astronomical Objects (TAOs) on real data

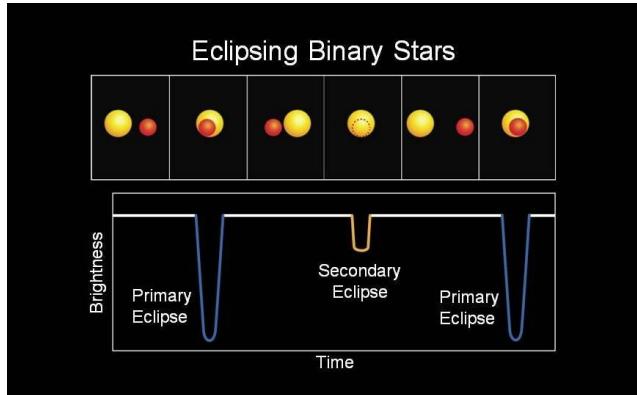
TAO Light curves

- Luminosity - "amount of light emitted"

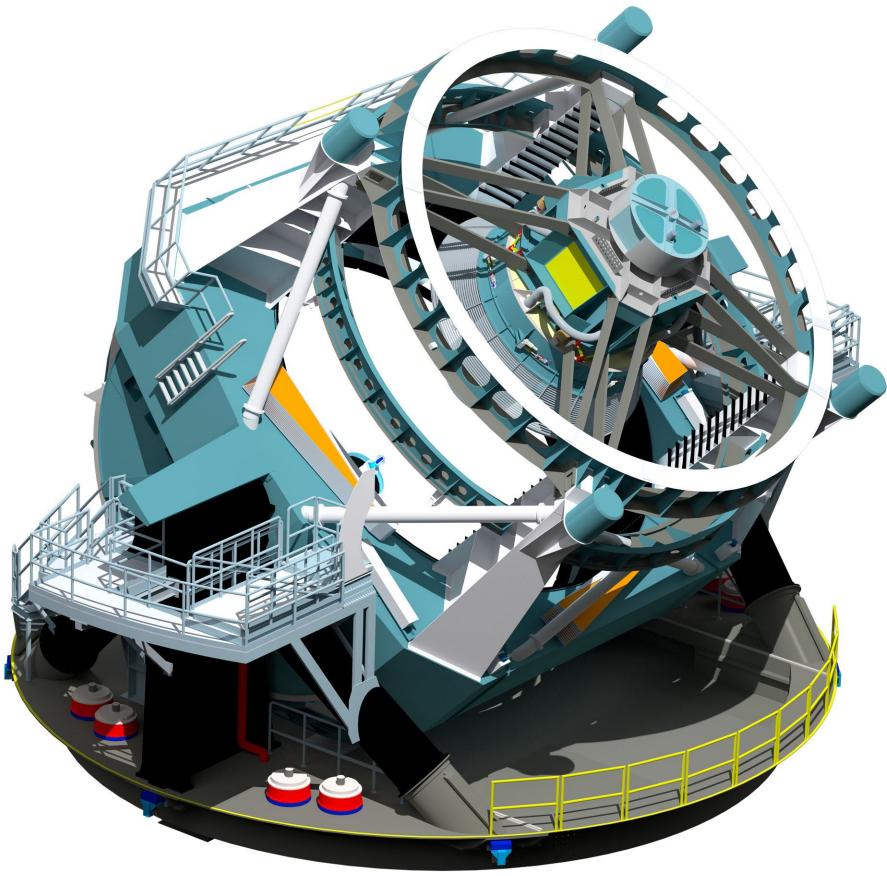
Non-periodic



TAO



Non - TAO



LSST

Large Synoptic Survey Telescope
Under construction, ready in 2022
3.2 GP



LSST Location

LSST = Open data

- 15 TB/day
- 10^6 TAOs/day

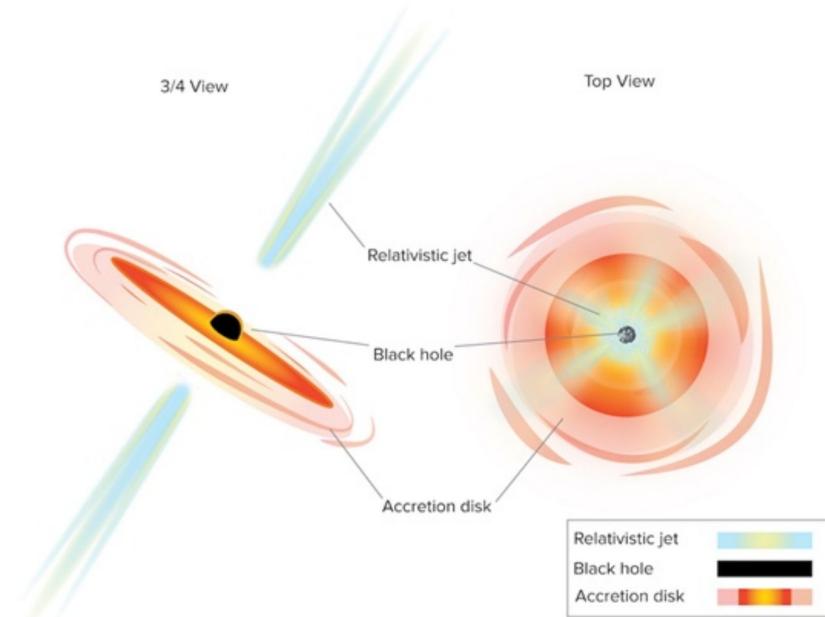


0.00004%

AGNs, BZs

Scientific importance

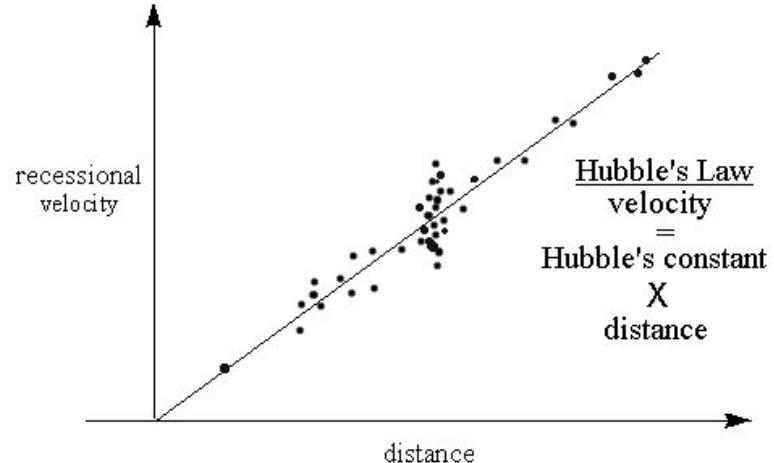
- At the center of galaxies
- Have black holes as nuclei



Supernovae (1a)

Scientific importance

- Hubble's constant
- Hubble's law
- Element formation
- Star life

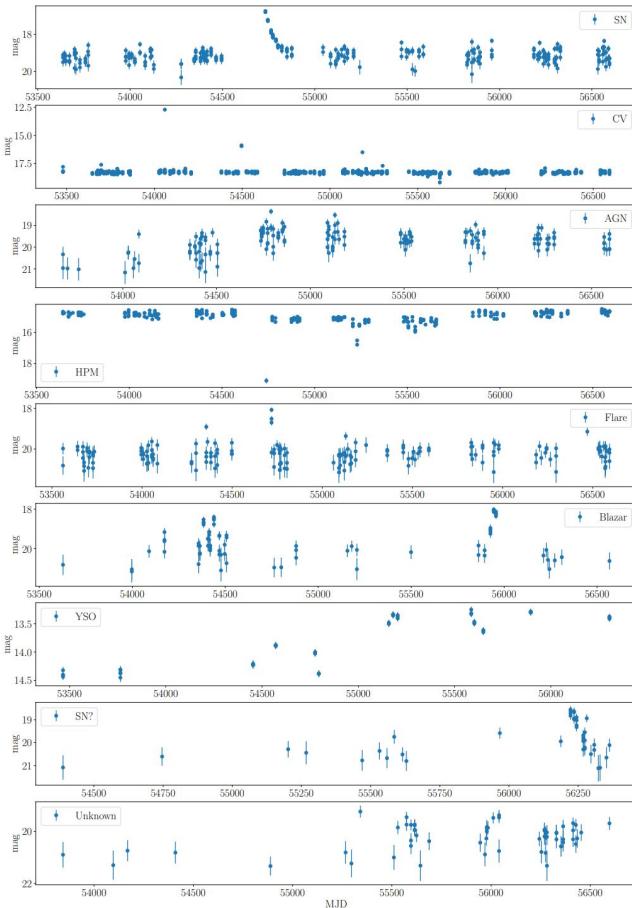


Data

Database

Catalina Real Time Survey (CRTS)

- Light curves
- Catalina Sky Survey
 - Mt. Lemmon Survey (MLS)
 - Catalina Sky Survey (CSS)
 - Siding Spring Survey (SSS).
- Variable length sequences
- Main Classes:
 - AGN
 - SN
 - CV
 - HPM
 - Blazar
 - Flare
 - Other
 - Non-TAO



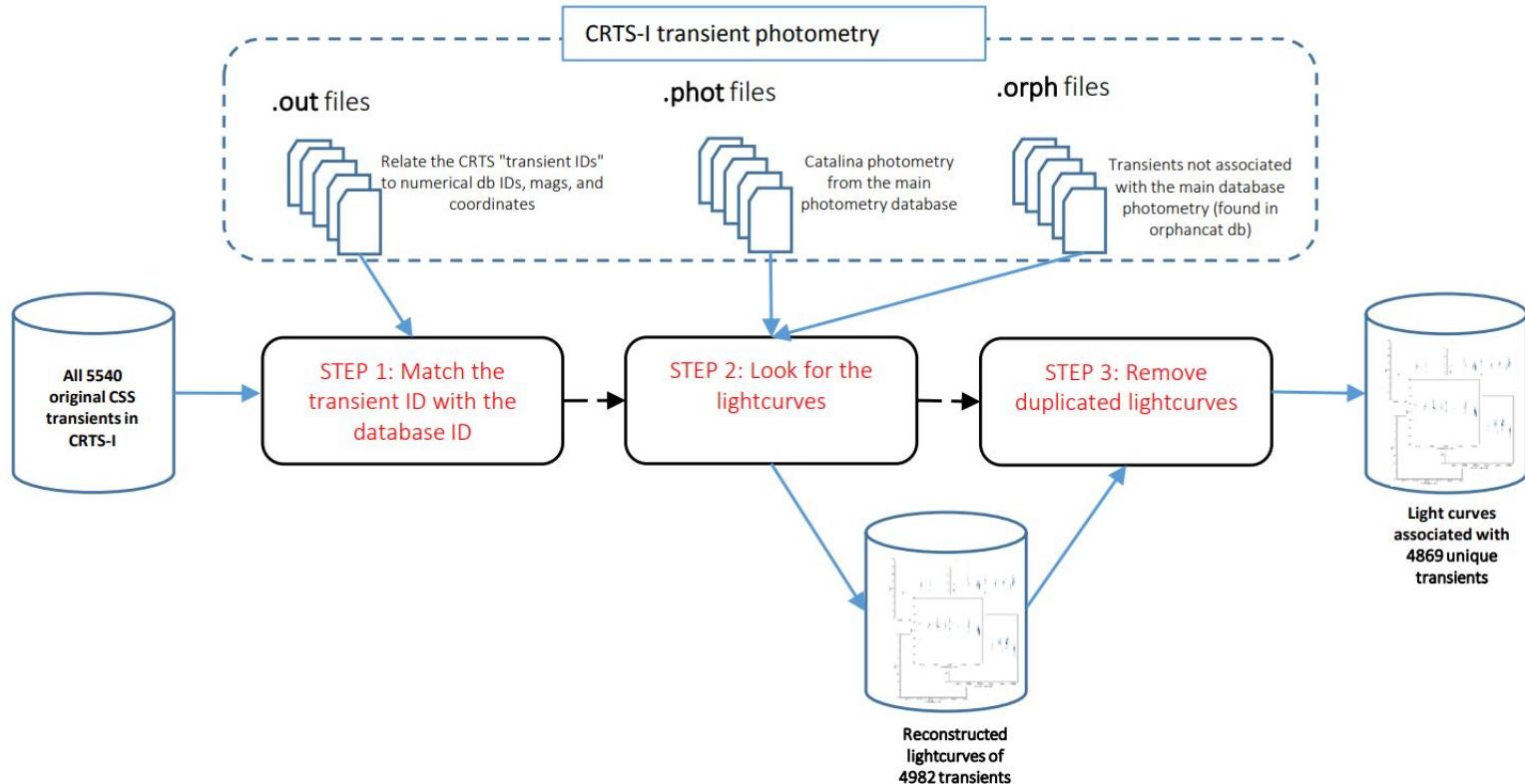


Figure 1. MANTRA Dataset Set Up: Lightcurve compilation for transient classes.

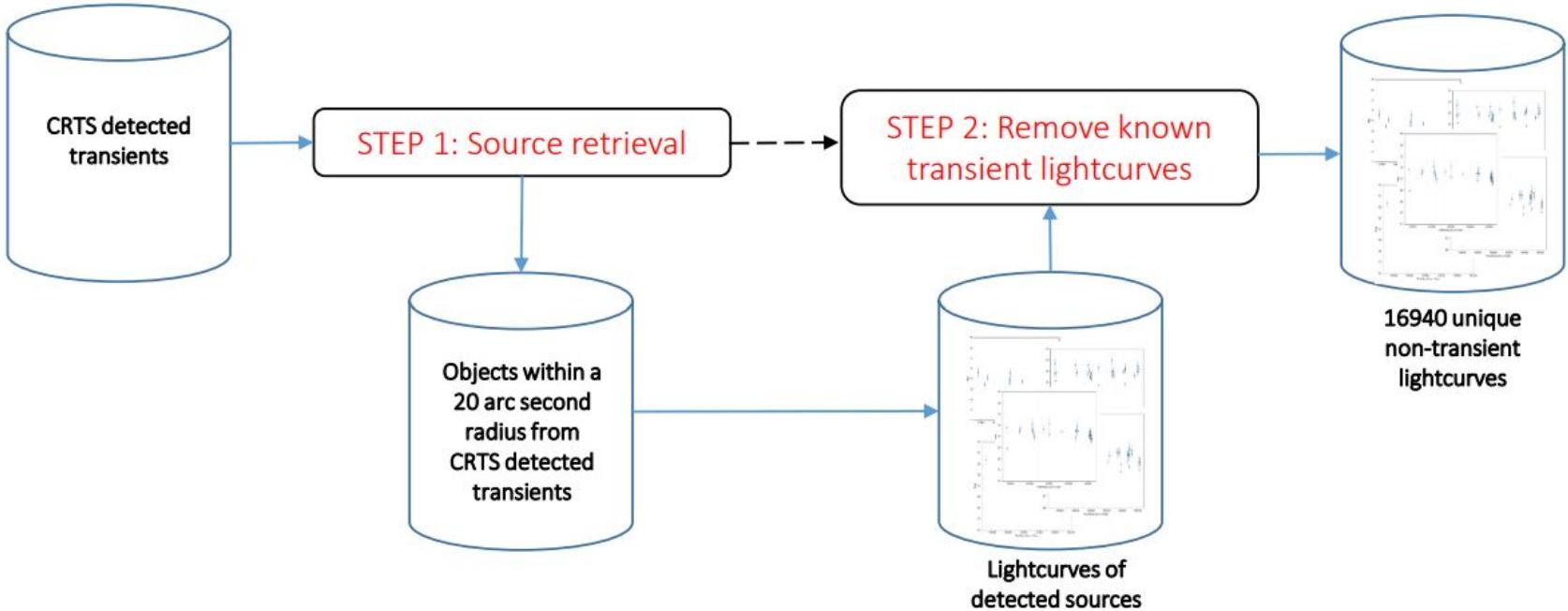
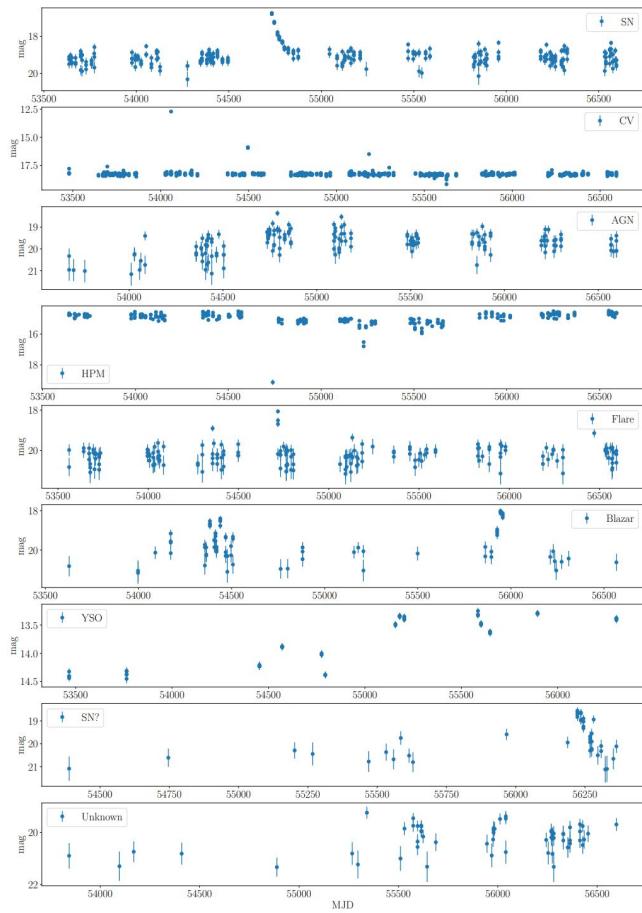
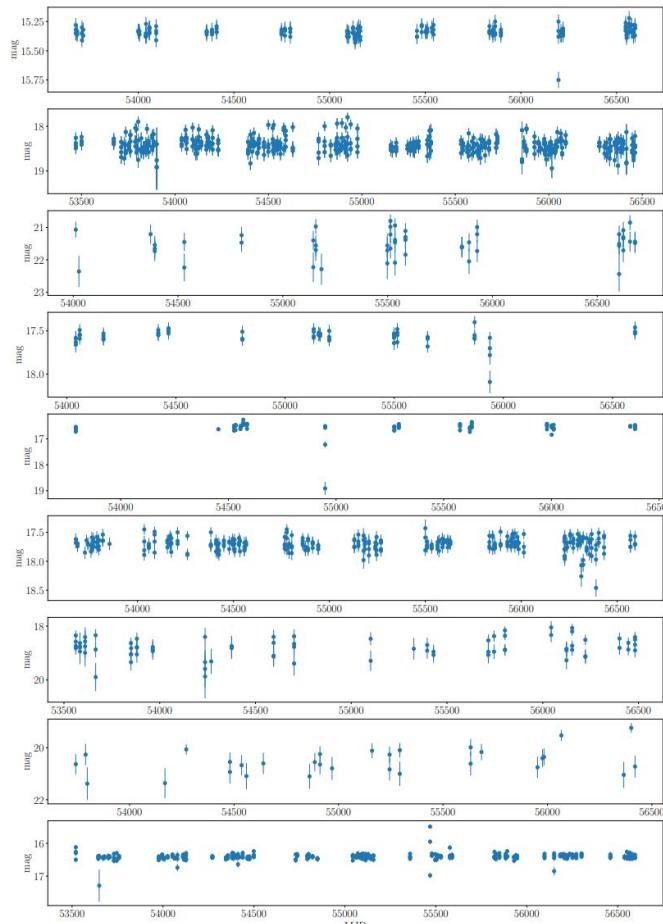


Figure 2. MANTRA Dataset Set Up: Lightcurve compilation for non-transients.



TAOs



Non-TAOs

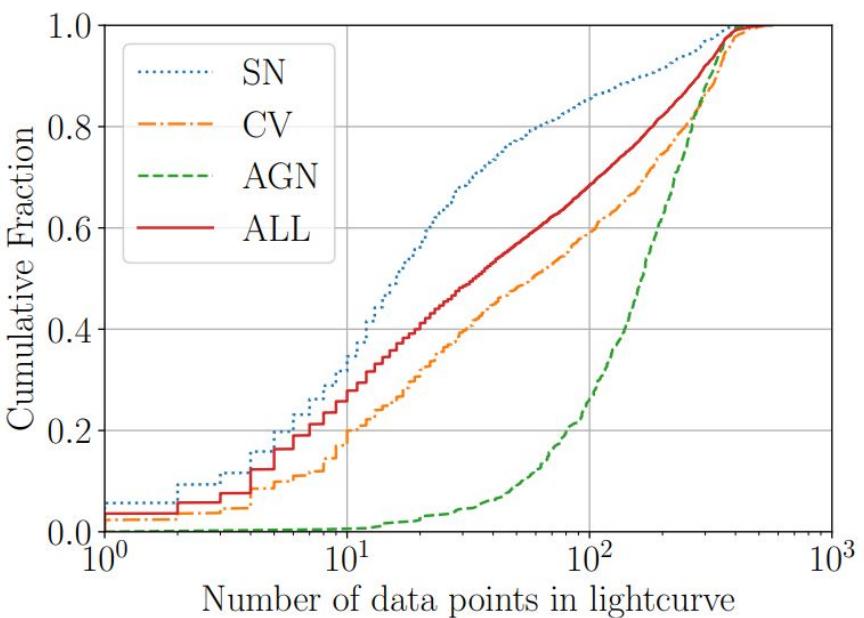
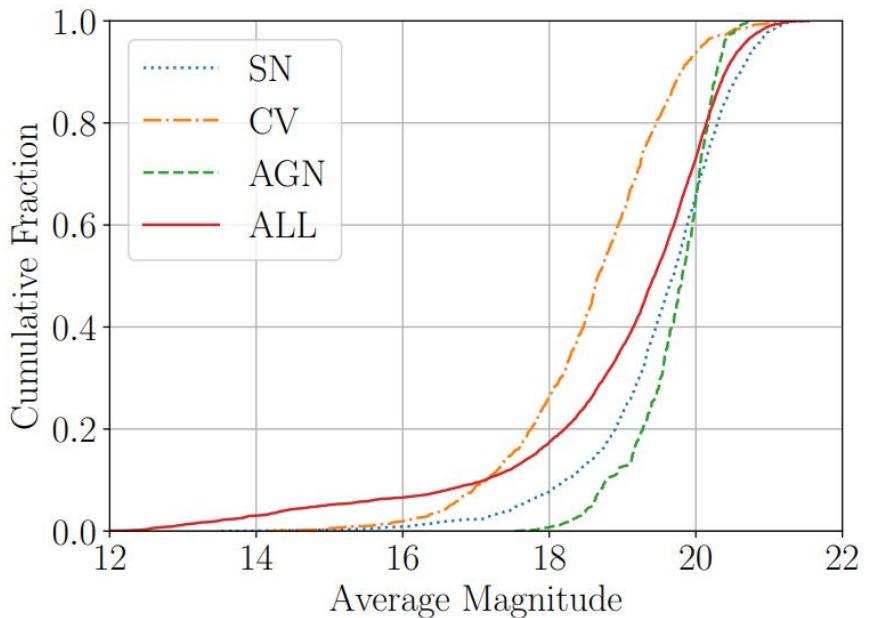


Figure 3. Cumulative number of lightcurves (expressed as a fraction) as a function of average magnitude (left) and number of data points in the lightcurve (right). This includes information for the three most representative classes (SN, CV, AGN) and the whole database (ALL).

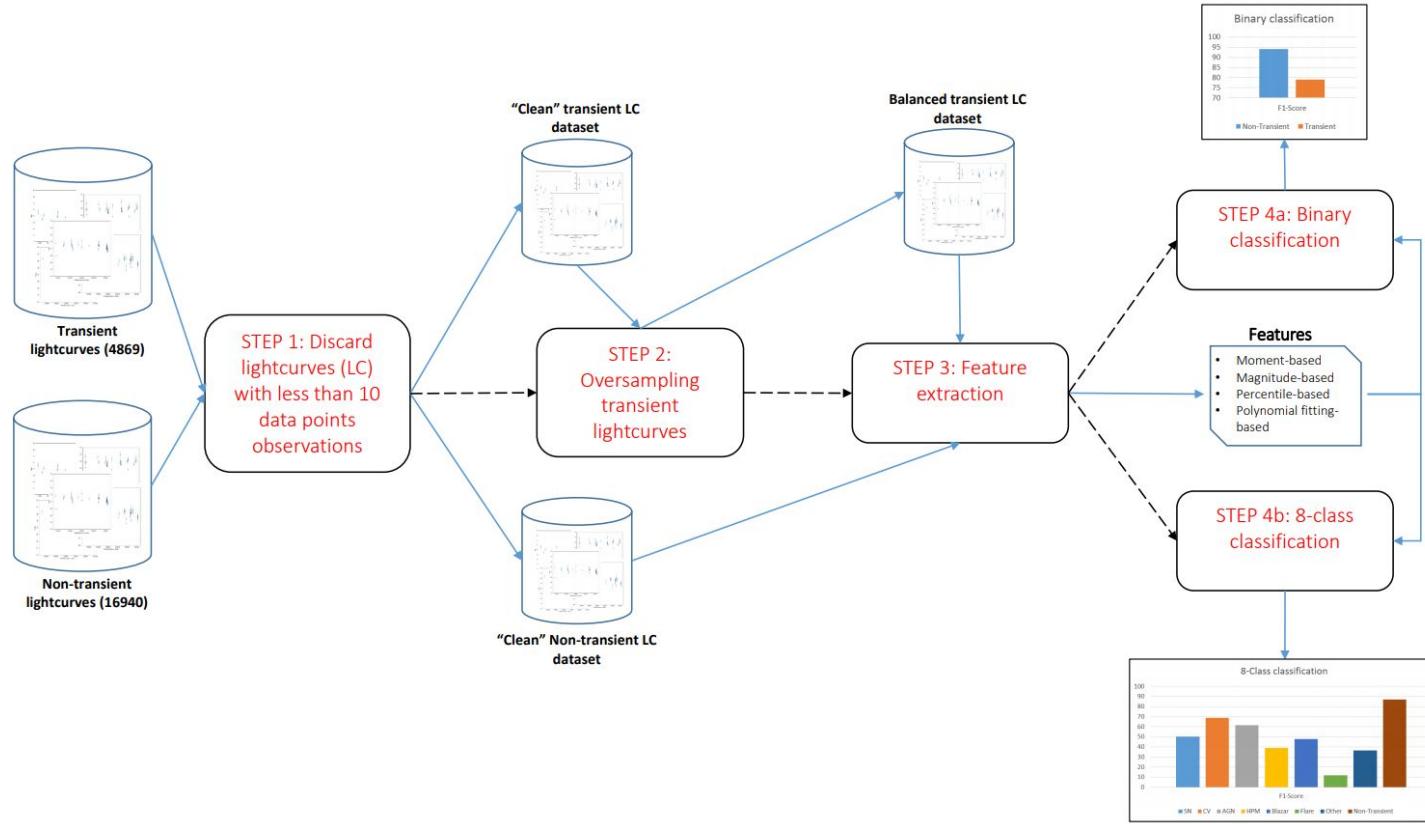


Figure 6. Overview of the Machine Learning process on the MANTRA dataset for the binary and 8-class classification tasks. We take the raw lightcurves as input, preprocess the data (step 1) and balance the classes for the training phase (step 2). We extract the features (step 3) that are feed into the ML algorithms that perform the classification task (step 4).

How do we classify
TAOs?

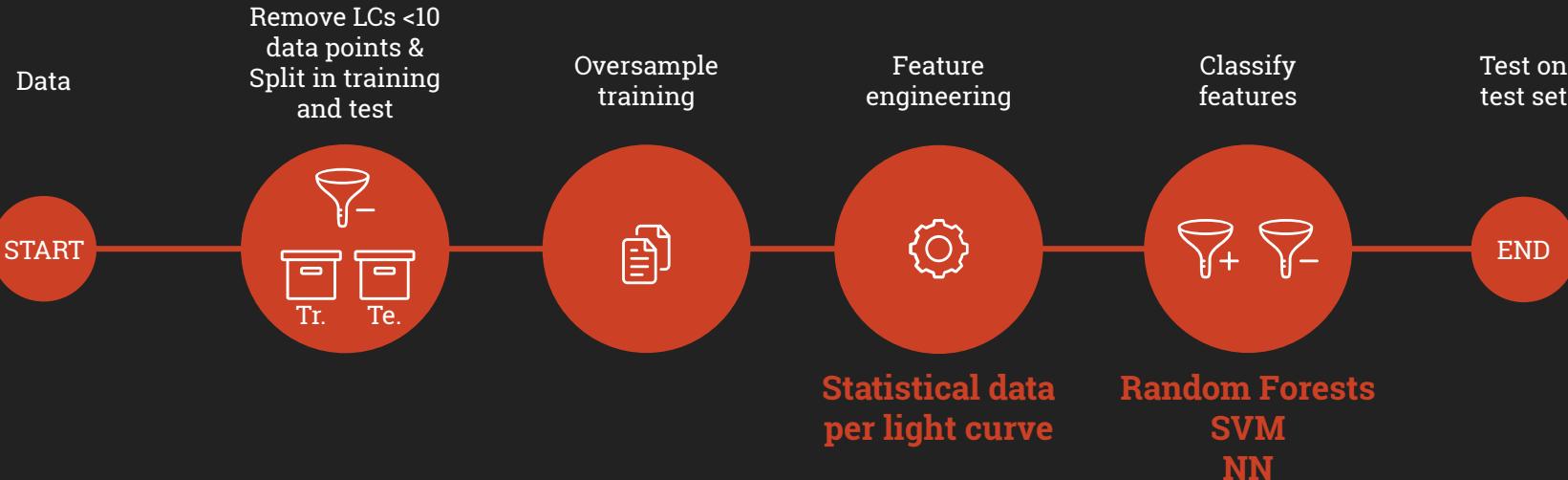


Light-curves

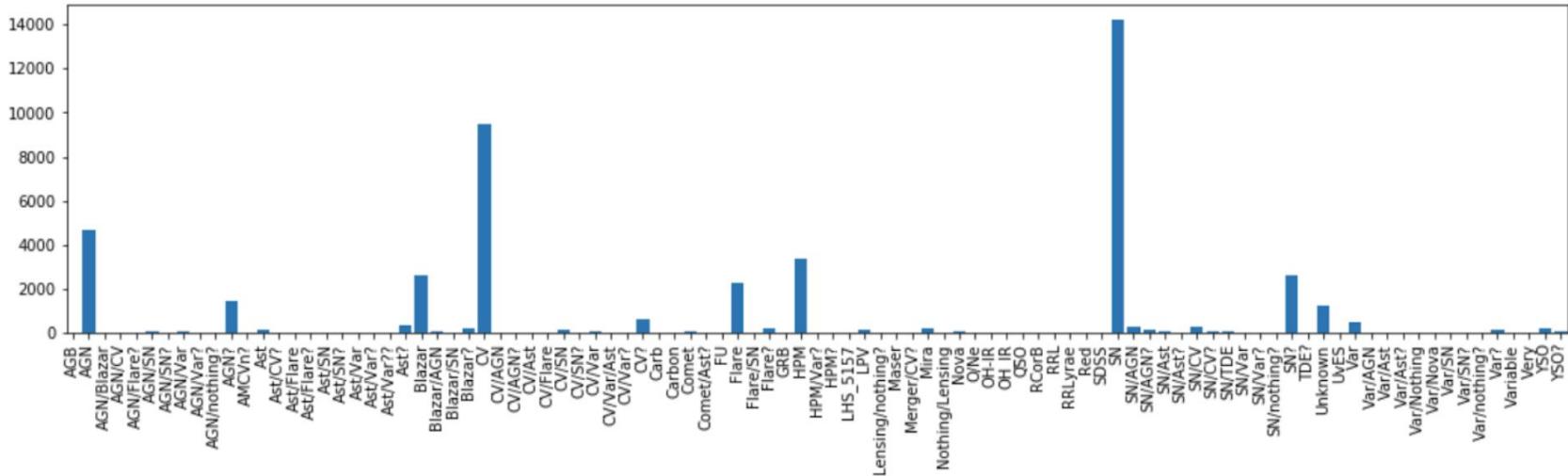
**Machine
learning**

Classification process

Diego's Pipeline - Transients



Data Oversampling



Features

31 total features per object

- Moment based features
 - Skew
 - Kurtosis
 - Small sample kurtosis
 - Standard deviation...
- Magnitude based features
 - The difference between the maximum and minimum magnitudes.
 - Maximum absolute slope between two consecutive observations.
 - The median of the difference between magnitudes and the median magnitude...
- Percentile based features
 - Largest percentage difference between the absolute maximum magnitude and the median.
 - Ratio between F5,95 and the median flux...
- Polynomial fitting features
 - Poly1_T1
 - Poly2_T1
 - Poly2_T2...
 - Poly4_T4

Results

General

Binary Classification:

- TAO
- Non-TAO

8 Class Classification:

- | | |
|-------|-----------|
| ○ AGN | ○ Blazar |
| ○ SN | ○ Flare |
| ○ CV | ○ Other |
| ○ HPM | ○ Non-TAO |

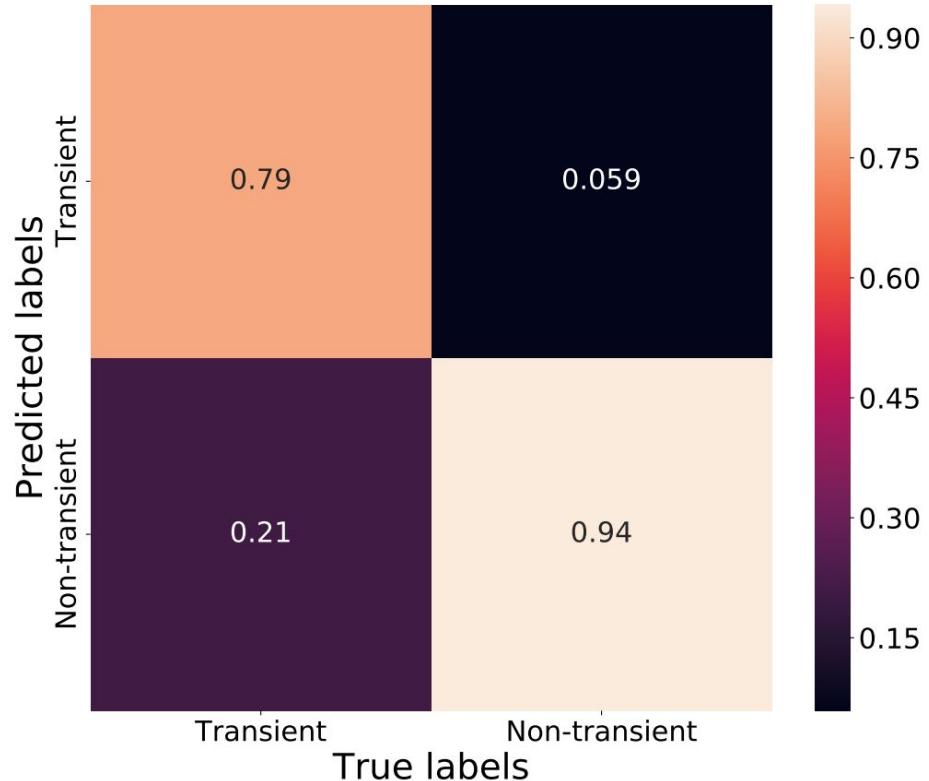
Case	Classifier	Precision	Recall	F1-score
Binary	RF	86.61	86.61	86.61
	SVM	82.61	82.525	82.57
	NN	71.84	73.19	72.51
8 Class	RF	46.25	63.59	50.38
	SVM	32.94	55.22	36.60
	NN	36.0	55.19	38.59

Table 2. Average precision, recall and F1-score across all classes for each algorithm and classification task. Best results per metric per classification task are in bold.

Results

Binary - Best

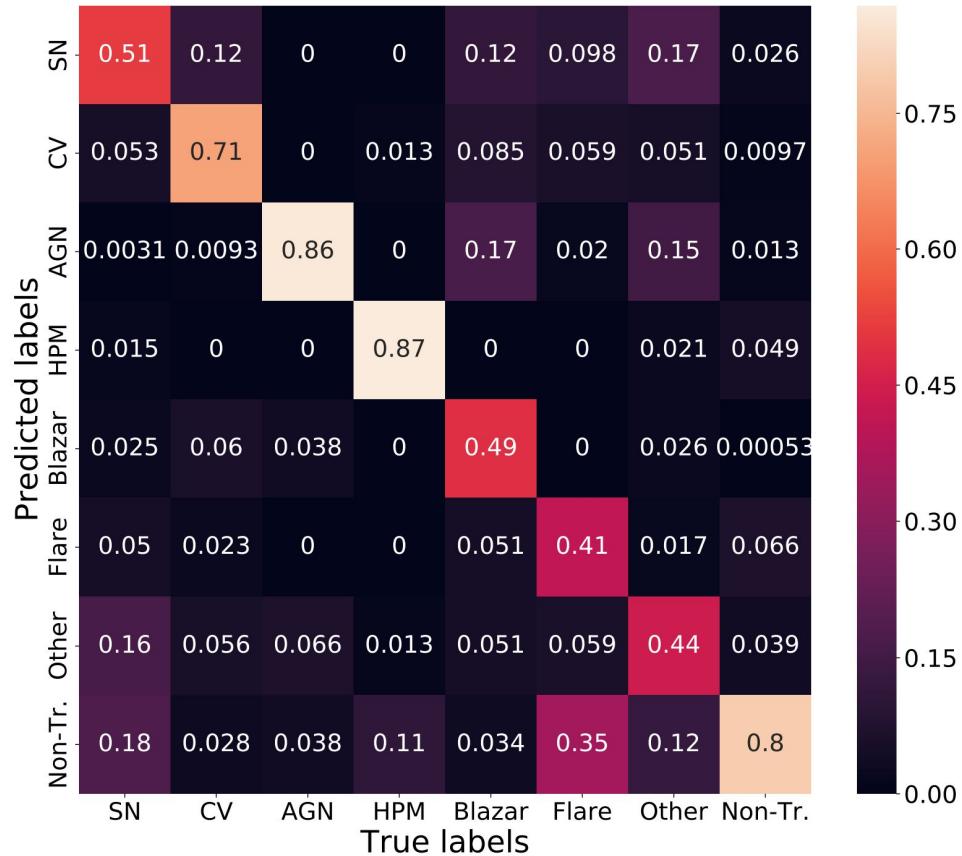
Random Forests



Results

8 Class - Best

Random Forests



Results

8 Class - Best

Random Forests

Class	Precision	Recall	F1-score	Cover
SN	48.82	51.39	50.07	323
CV	66.96	70.69	68.77	215
AGN	48.14	85.84	61.69	106
HPM	25.19	86.84	39.05	76
Blazar	46.77	49.15	47.93	59
Flare	7.00	41.17	11.96	51
Other	31.11	44.01	36.46	234
Non-Tr.	96.06	79.69	87.12	3798
avg/total	46.25	63.60	50.38	4862

Table 3. Precision, Recall and F1-Score for the 8-Class Classification Task.

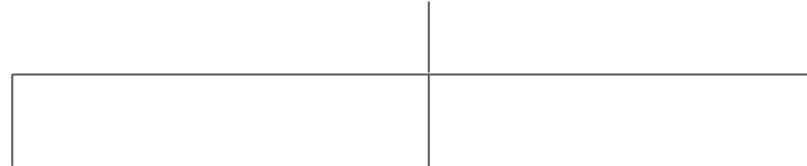
Conclusions

- Data compilation based on data from the Catalina Real-Time Transient Survey.
 - 4869 transient
 - 16940 non-transients
- The dataset is publicly available at <https://github.com/MachineLearningUniandes/MANTRA>
- **New benchmark** for TAO classification from **REAL light curves**

What's next?

- Deep Learning
- Images

Recurrent Neural Network



LSTM Transformers GRU

Thank you