Tuning MicroLIA for Rubin LSST microlensing light curve identification

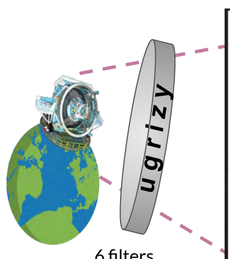
K. Nowogrodzki [1], A. Varela [1], M. Makler[1,2], D. Godines[3], R. A. Street[3]

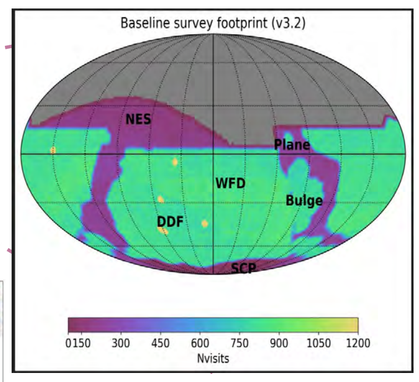
& the TVSSC microlensing sub-group.

[1]ICAS/ICIFI, ECyT/UNSAM & CONICET, [2]CBPF, [3]Las cumbres observatory.

Context

Rubin Observatory will carry out a wide-field imaging survey in the time domain, enabling, among many other applications the first "all sky" microlensing survey, with thousands of expected events, from planetary systems to massive black holes, opening a new discovery space.





Simulated number of observations in 10 years of 20.000 deg^2 of the entire Southern Hemisphere visible sky (1)

- **high depth** and **image quality** with cadence of a **few days**
- **millions** of transient alerts every night
- Thousands microlensing events detection

Objectives

Improve the detectability of **microlensing events among numerous variable sources:**

- Improve a **method for classification** of variable sources.
- Study the impact of the various proposed survey strategies
- Determine the efficiency and contamination of the detection of microlensing events
- Explore the effects of combination with other surveys (e.g. Roman Space Telescope)

Prepare for the use of brokers to classify microlensing events and trigger follow-ups

MicroLIA: software to train machine learning classifiers for microlensing. It computes light curves's properties about their symmetry, statistics and more. Then, with this software you can train models like Random Forest, eXtreme Gradient Boosting, Neural Network or Convolucional Neural Network.

Methodology

The workflow is divided in 4 main steps: **constant light curve simulations** with LSST cadence and errors (currently for the galactic bulge, in the *i* band), testing of different selection (**cuts**) schemes for the data, **training and testing** a light curve classifier method for microlensing (machine learning models) and **simulating light curves for all of confusable categories with microlensing.**.

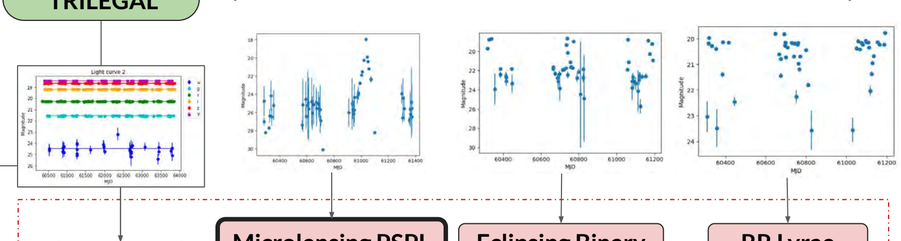
Light curves simulation

Observation strategy: **OpSim**

Sources stars of galactic center simulation: **TRILEGAL**

ELAsTiCC challenge

Light curve simulation with photometric errors and cadence



Constant | **Microlensing PSPL** | Eclipsing Binary | RR Lyrae | dwarf-nova | δ-Scuti | Mdwarf-flare | Cepheid

Cuts

Apply cuts

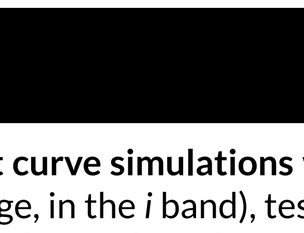
Cut: "Valid" — *Well measured* data points with at least 10 points

cut_m_sat
Delete points with $m_i > m_{sat}$

cut_m_5
Has > 10 points correctly measured
$m_i + \sigma_i < m_{5\sigma}$

"Peak" — **has_peak**
At least 4 points 3 σ brighter than the median
$m_i + 3\sigma_i < median$

"Chi" — **chi2_dof_cut**
Limit the X^2/dof as a deviation from constant
$$\chi^2_{dof} \equiv \frac{\sum_{i=1}^{N}\left(\frac{m_i - median}{\sigma_i}\right)^2}{N}$$

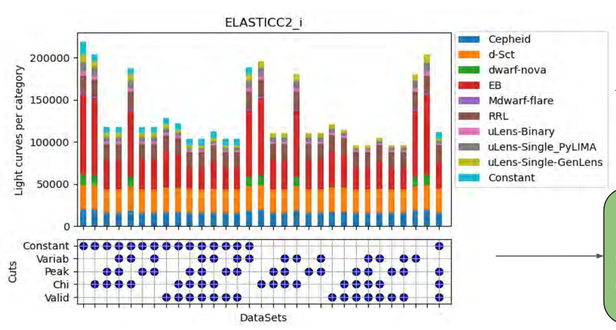


$m_{err} = \sigma$

dof = degrees of freedom

N = number of data points for the light curve

Baseline model: First train for several combination of cuts with just 4 categories test performance. Best combination: Valid_Constant_Chi



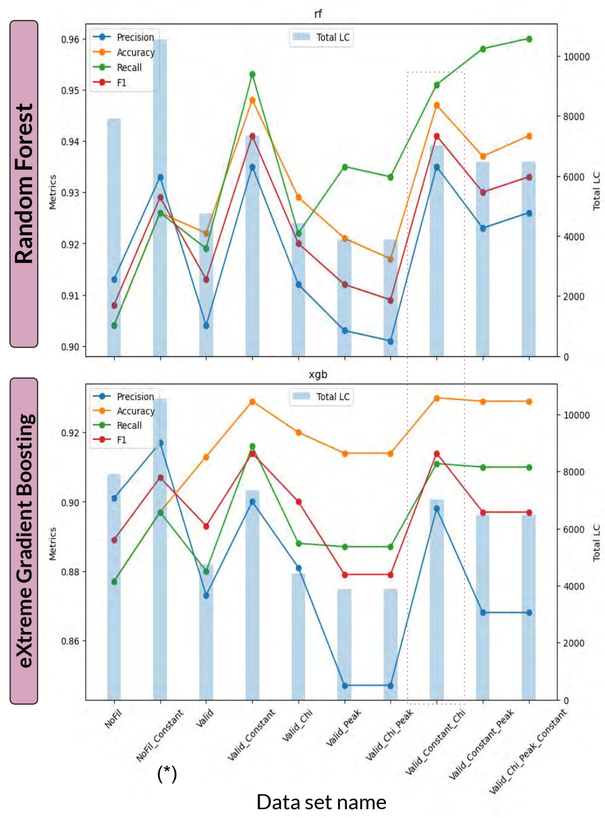
Second train with Valid_Constant_Chi → Classification models with **MicroLIA**
- RandomForest
- eXtreme Gradient Boosting

Problems:
- none Mdwarf-flare
- <1000 dwarf-nova

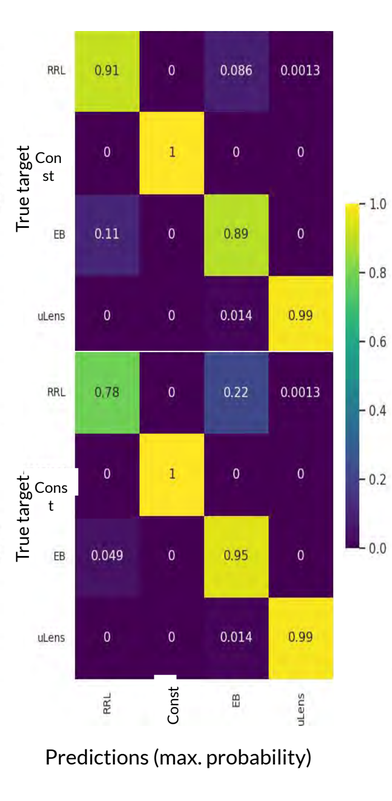
→ Simulation of more light curves

Preliminary results

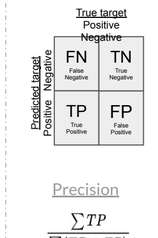
Score metrics and total amount of light curves used in the data set (train and test o all classes).



Confusion matrix of Valid_Constant_Chi



Score metrics



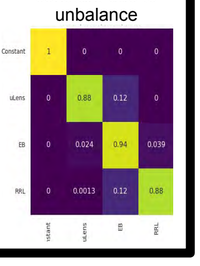
Precision $\frac{\sum TP}{\sum(TP+FP)}$

Accuracy $\frac{\sum(TP+TN)}{\sum(TP+FP+FN+FN)}$

Recall $\frac{\sum TP}{\sum(TP+FN)}$

F1 score $\frac{2\,.\,precision\,.\,recall}{\sum(precision+recall)}$

(*) Optimized for constant due to unbalance



Conclusions

Impact of cuts: Optimal performance comes from Valid_Constant_Chi, providing constant class for classification, well measured points and detectable variability.

Classification of transients and variable events:

Explored impact of:
- Data selection:
  - Filtering well-measured curves ("Valid")
  - Imposing variability ("Peak", "Chi")
- Introduction of a "constant" category (which was not present in ELAsTiCC dataset) → Improved metrics.

Simulation of light curves

Demonstrated the ability to generate limitless data that accurately replicates LSST's cadence and errors:
- Microlensing light curves
- Constant light curves

Future steps

This work is in process and we hope obtain information that allow us to determine a good strategy for the data combination and improve the classification of microlensing events. Here we list a few further steps:

- **Simulating BE stars, Cataclysmic Variables and MIRA with realistic cadence and error bars to training a model and integrate it to FINK**
- **Train with larger datasets with the simulated light curves.**
- **Multiband classifier: Add a new color feature to train with MicroLIA**
- Train on different locations in the sky.
- Analyze MicroLIA hyperparameters and features.
- Add other kinds of transient phenomena to the datasets.
- Test different observation strategies for Rubin

References

D. Godines, E. Bachelet, G. Narayan and R.A. Street, *A machine learning classifier for microlensing in wide-field surveys*, arXiv:2004.14347 [astro-ph.IM].

S. Sajadian, R. Poleski, *Prediction on detection and characterization of Galactic disk microlensing events by LSST*, arXiv:1806.06372 [astro-ph.IM].

R. Kessler et al. *Models and Simulations for the Photometric LSST Astronomical Time Series Classification Challenge (PLAsTiCC)*, DOI 10.1088/1538-3873/ab26f1