Microlensing signatures of extended dark objects using machine learning

BeyondWIMPS
Durham University
March 2024

Miguel Crispim Romão
IPPP
Durham University
miguel.romao@durham.ac.uk

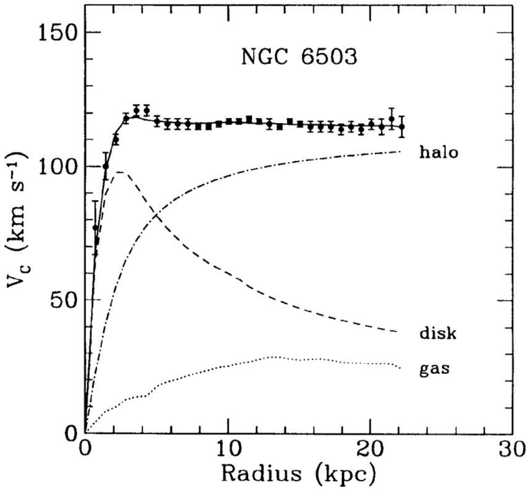
Outline

- Introduction
  - Motivation
  - Microlensing of Compact Dark Objects
- Machine Learning Signatures of Dark Objects
  - OGLE2 Adaptive Cadence
  - Regular Daily Cadence
- Current Work
- Conclusions and Outlook

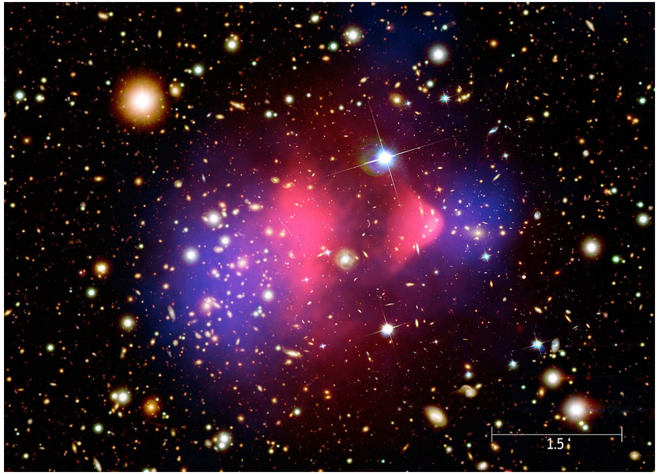
Introduction

Introduction
Motivation

- Plenty and diverse evidence for Dark Matter
    - Astrophysical



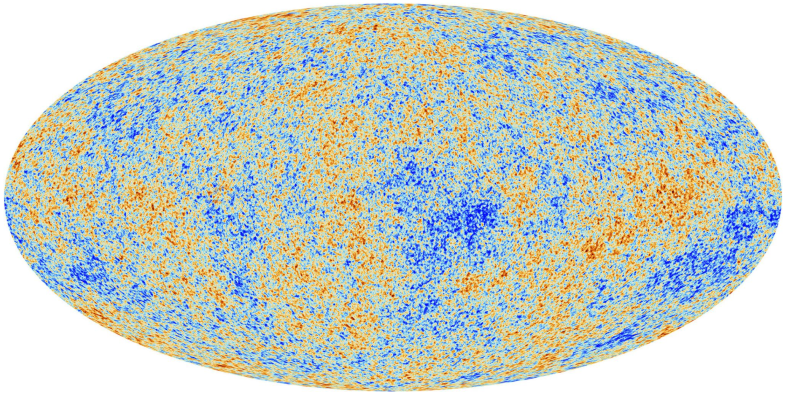
dx.doi.org/10.1093/mnras/249.3.523



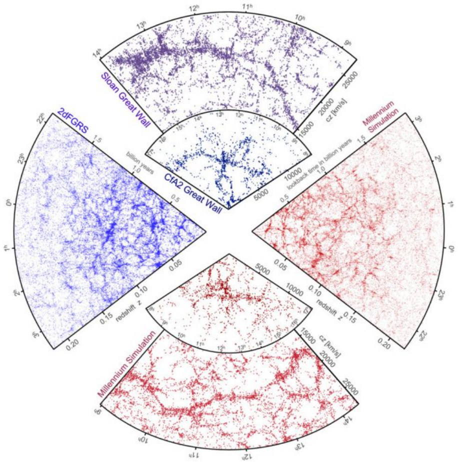
dx.doi.org/10.1093/mnras/249.3.523

Introduction
Motivation

- Plenty and diverse evidence for Dark Matter
    - Cosmological



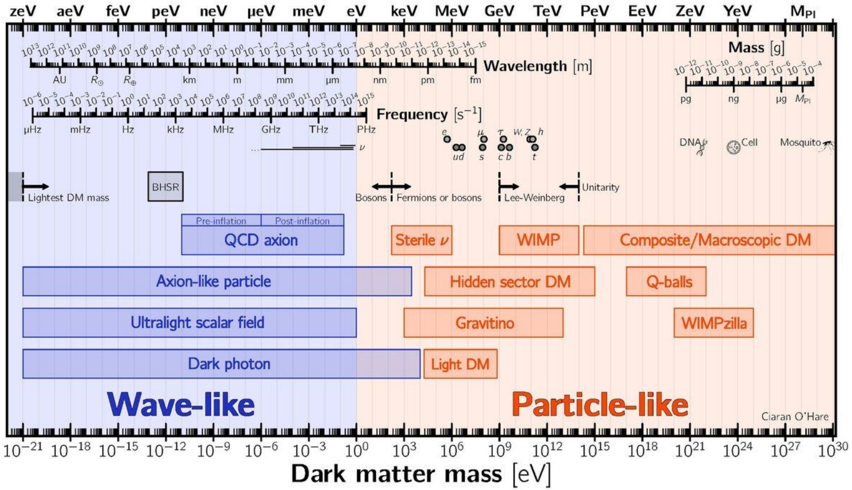
ESA



astro-ph/0604561, doi:10.1038/nature04805

Introduction
Motivation

All evidence is gravitational. Nonetheless, the dominant paradigm has been field theoretical.



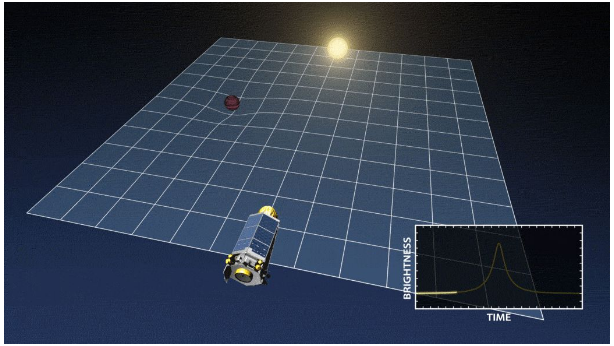
Wikipedia

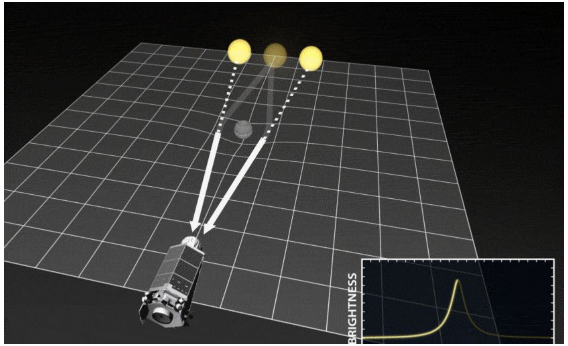
What about here?

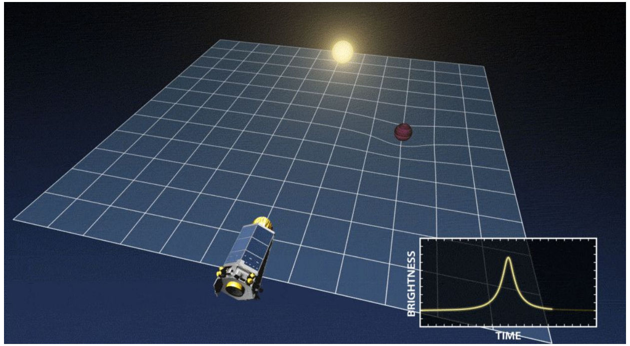
Dark celestial objects that can be detectable/constrained using microlensing!

Introduction
Microlensing of Compact Dark Objects

- Matter bends space, changing geodesics traveled by light
- Different from strong lensing as images do not resolve
  - Instead an increase in brightness is obserserved



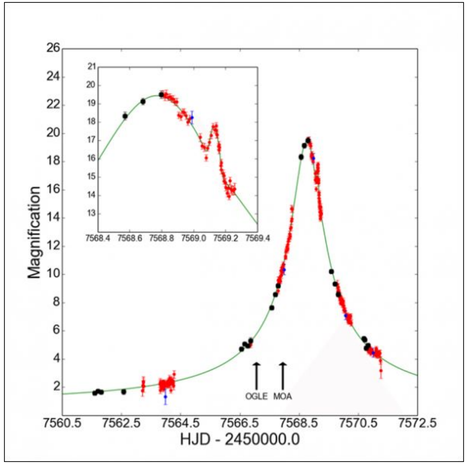




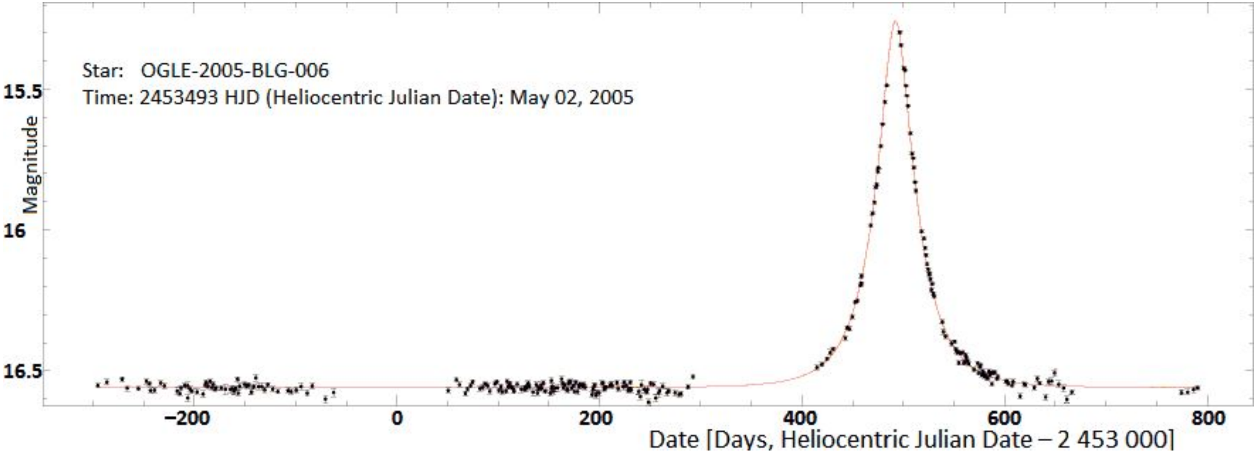
Wikipedia

Introduction
Microlensing of Compact Dark Objects

Depending on the nature of the lens, different brightness magnification profiles emerge…



British Astronomical Association



OGLE collaboration

Introduction
Microlensing of Compact Dark Objects

Let's define the source-lens-observer geometry and associated quantities:

- β: Impact angle (i.e. lens position)
- θ: Image angles
- ξ: Impact parameter

and the normalised quantities

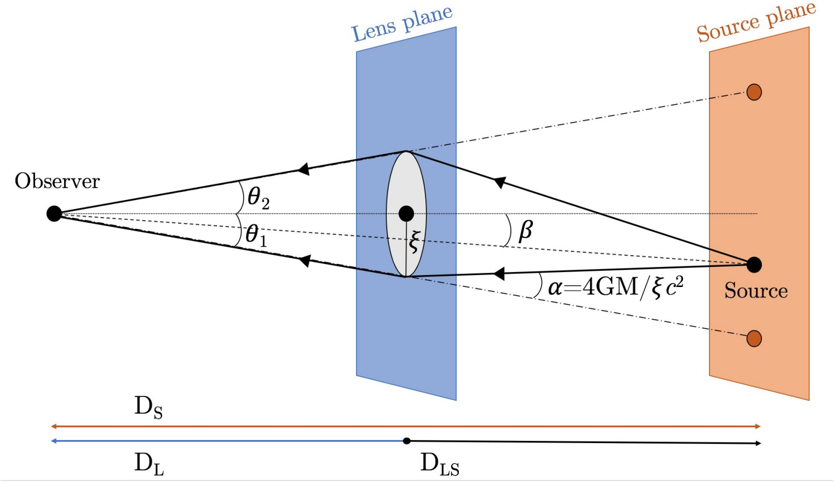
- $u = \xi/r_E = \beta/\theta_E$
- $\tau = \theta/\theta_E$
- $x = D_L/D_S$

Where

$$\theta_{\rm E} \equiv \sqrt{\frac{4GM}{c^2}\frac{D_{\rm LS}}{D_{\rm L}D_{\rm S}}}$$

is the Einstein angle for a point lens at β=0 and $r_E=D_L\theta_E$ the associated Einstein radius.

Croon, McKeen, Raj 2020



Introduction
Microlensing of Compact Dark Objects

The position of the images, τ_i, are given by the solutions of the lens equation

$$u = \tau - \frac{m(\tau)}{\tau}$$

where m(τ) is the mass profile of the lens, which accounts for the distribution of the lens mass projected onto the lens plane

$$m(\tau) = \frac{\int_0^\tau d\sigma\sigma \int_0^\infty d\lambda\, \rho(r_{\mathrm{E}}\sqrt{\sigma^2 + \lambda^2})}{\int_0^\infty d\gamma\gamma^2 \rho(r_{\mathrm{E}}\gamma)}$$

The total magnification, μ, will be given by the sum of the magnifications of each of each image

$$\mu = \sum_i \left| \frac{\theta_i}{\beta} \frac{d\theta_i}{d\beta} \right| = \sum_i \left| \frac{\tau_i}{u} \frac{d\tau_i}{du} \right|$$
$$= \sum_i \left| 1 - \frac{m(\tau_i)}{\tau_i^2} \right|^{-1} \left| 1 + \frac{m(\tau_i)}{\tau_i^2} - \frac{1}{\tau_i} \frac{dm(\tau_i)}{d\tau_i} \right|^{-1}$$

Introduction
Microlensing of Compact Dark Objects

For a point-like lens, such as a Primordial Black-Hole (PBH) we have m(τ)=1 and the lens equation has two solutions, leading to

$$\mu_{\rm tot} = \frac{u^2+2}{u\sqrt{u^2+4}}$$

Therefore, when the lens “crosses” its Einstein radius, u=1, we can observe a magnification μ=1.34.

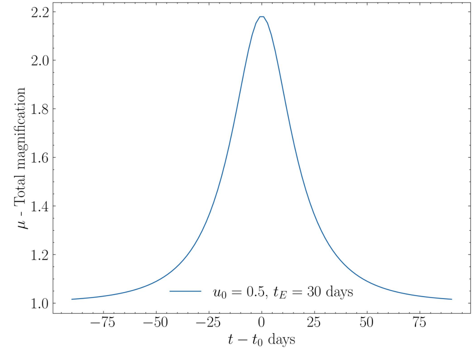
To connect to surveys, we notice that while traversing its plane, the lens impact parameter is related to the survey time, t, as

$$u = \sqrt{u_0^2 + \left(\frac{t-t_0}{t_E}\right)^2}$$

where u_0 is the lens minimal (normalised) impact parameter, t_E is the lens “Einstein crossing time”, i.e. the time it takes for the lens to cross its Einstein radius, and t_0 the time of the magnification peak.

Introduction
Microlensing of Compact Dark Objects

One can then easily simulate the magnification of a point-like lens.



Surveys look for light curves with this shape and count the number of matching events. This can be used to constraint the **fraction** of Dark Matter composed by PBH:

$v_{\mathrm{E}}(x) \equiv 2u_{1.34}(x)r_{\mathrm{E}}(x)/t_{\mathrm{E}}$ ($u_{1.34}$ is such that $\mu(u \leq u_{1.34}) \geq 1.34$, $u_{1.34}=1$ for point-like)

$$\frac{d^2\Gamma}{dx\,dt_{\mathrm{E}}} = \varepsilon(t_{\mathrm{E}})\frac{2D_{\mathrm{S}}}{v_0^2 M}\boxed{f_{\mathrm{DM}}}\rho_{\mathrm{DM}}(x)v_{\mathrm{E}}^4(x)e^{-v_{\mathrm{E}}^2(x)/v_0^2}$$

Survey efficiency

220 km/s

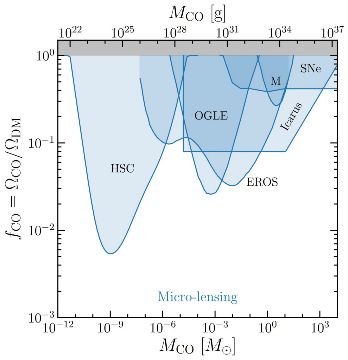
Halo profile (isothermal)

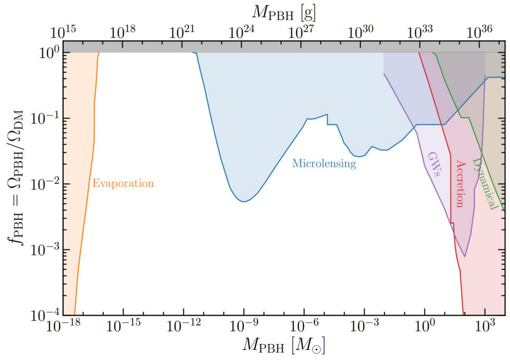
$$N_{\mathrm{events}} = N_\star T_{\mathrm{obs}}\int_0^1 dx \int_{t_{\mathrm{E,min}}}^{t_{\mathrm{E,max}}} dt_{\mathrm{E}} \frac{d^2\Gamma}{dx\,dt_{\mathrm{E}}}$$

Number of observed stars

Total observation time

Min and maximal observable crossing times





Green, Kavanagh 2020

Machine Learning Signatures of Dark Objects

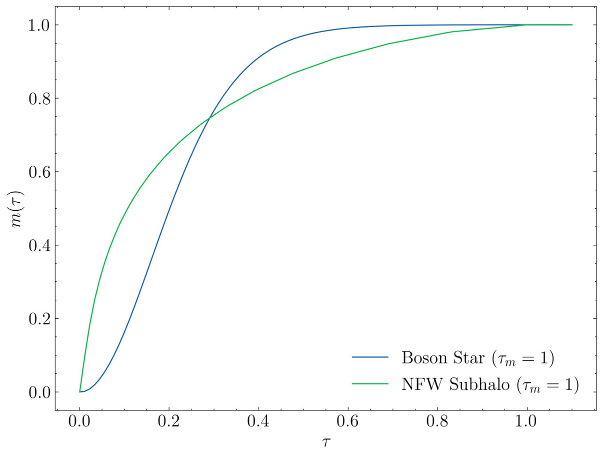
Microlensing signatures of extended dark objects using machine learning
Miguel Crispim Romão (IPPP, Durham), Djuna Croon (IPPP, Durham)
https://arxiv.org/abs/2402.00107 (submitted to PRD)

Machine Learning Signatures of Dark Objects
Extended Dark Objects

- So far we have only discussed point-like dark lenses, m(τ)=1
- However, many dark objects not only have substructure, but have extended structures and can be markedly not point-like
- Croon, McKeen, Raj 2020 and 2020, studied extended dark objects and constrained their population using microlensing data from EROS-2 (Magellanic Clouds) and OGLE-IV (Galatic Bulge) surveys
  - Two types of objects are of special interest for the work presented herein:
    - Boson Stars: gravitationally stable structures composed of scalar fields. Exhibiting a more disperse m(τ)
    - Navarro-Frenk-White (NFW) subhalos: products of hierarchical clustering of cold DM. Exhibiting a more peaked m(τ)
  - For sufficiently **flat** density profiles, the **caustics** empact the constraints

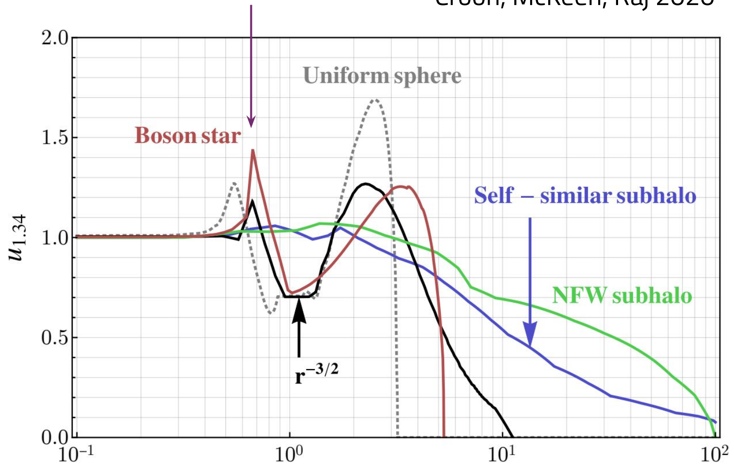
Machine Learning Signatures of Dark Objects

Extended Dark Objects



Caustics can emerge!

Croon, McKeen, Raj 2020

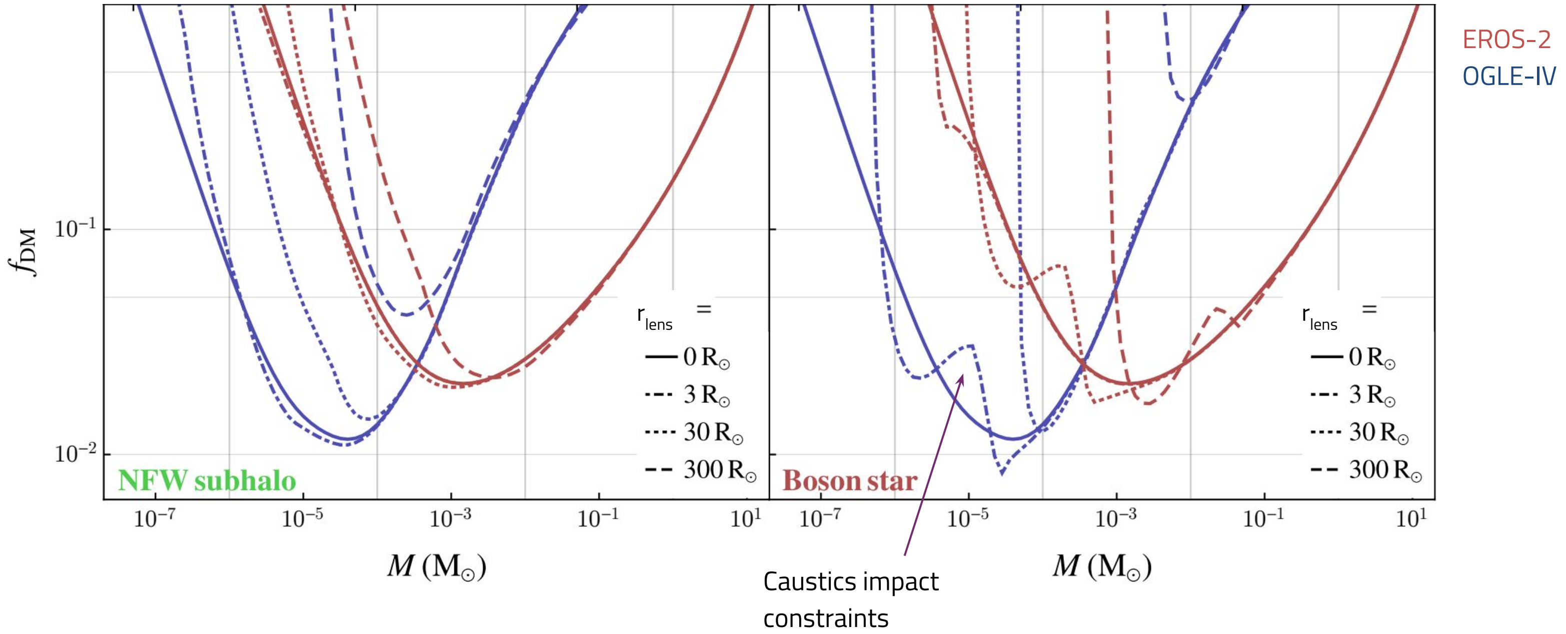


$\tau_m(x)$

Where $\tau_m = r_{lens}/r_E = \theta_{lens}/\theta_E$

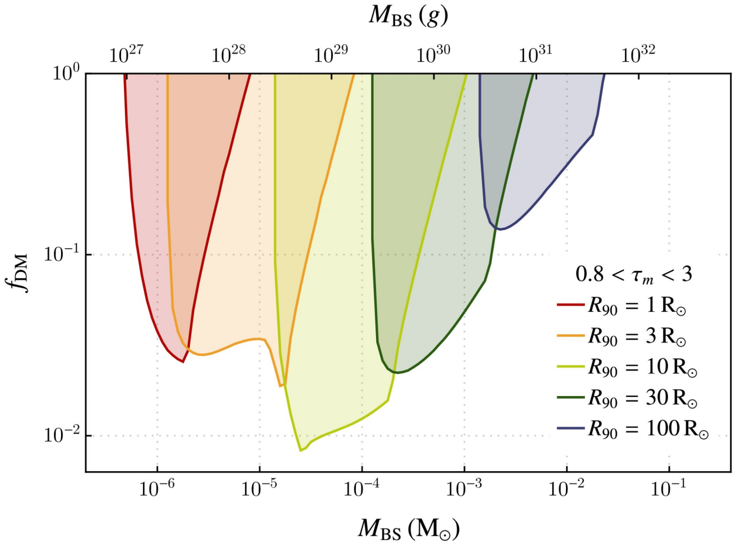
Machine Learning Signatures of Dark Objects
Extended Dark Objects

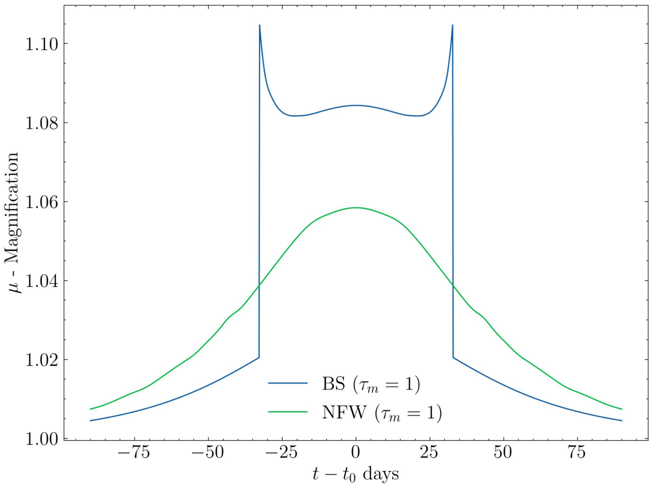
Croon, McKeen, Raj 2020



Machine Learning Signatures of Dark Objects
Extended Dark Objects

Besides constraining the fraction of DM composed of these objects, could we detect them?



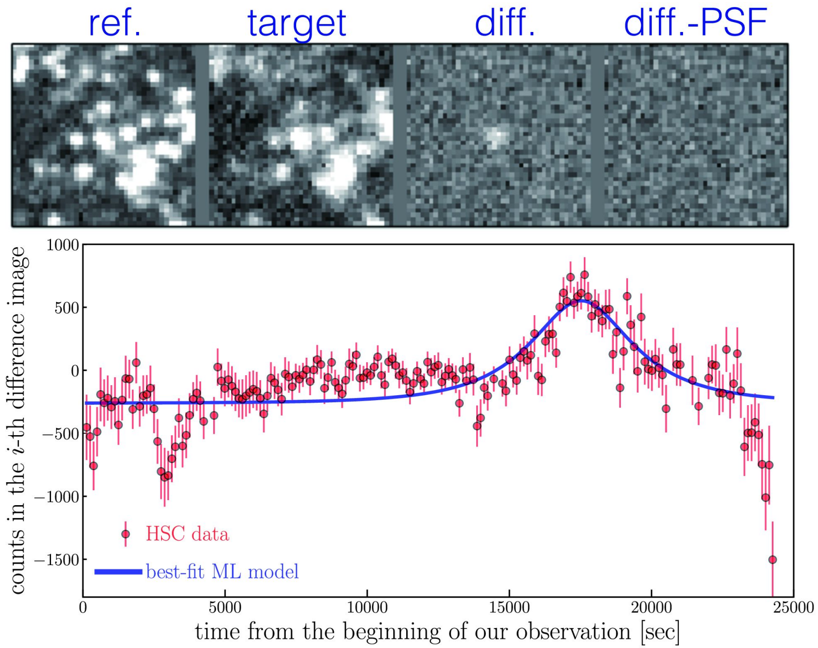


Here are already anticipating that the caustics will play a role for $0.8 \leq \tau_m \leq 3$

Machine Learning Signatures of Dark Objects
Machine Learning Microlensing

How are microlensing events detected, then?

- Images of a source are captured at different times with a certain **cadence**
- A reference image sets a **baseline** magnitude/flux for the source
- Differences between an image and the reference allow to see the **changes** in magnitude/flux
  - The **light curve** is the **time series** of the magnitude/flux over time



Niikura, H, et al - Subaru/HSC observation

Machine Learning Signatures of Dark Objects

Machine Learning Microlensing

- Data is not readily available
  - Only analysed and human-annotated events are released
- Analyses have a very constraining pre-analysis cuts focused on point-like microlensing signatures
- Nonetheless, we can perform a phenomenological study
  - **Simulate** light curves for multiple astrophysical phenomena
  - Use quantities derived from the time series to perform a **classification** task

Table 1. Selection Criteria for High-quality Microlensing Events in OGLE GVS Fields.

Criteria	Remarks	Number
All stars in databases		1,856,529,265
$\chi^2_{\mathrm{out}}/\mathrm{dof} \leq 2.0$	No variability outside a window centered on the event (duration of the window depends on the field)	
$n_{\mathrm{DIA}} \geq 3$	Centroid of the additional flux coincides with the source star centroid	
$\chi_{3+} = \sum_i (F_i - F_{\mathrm{base}})/\sigma_i \geq 32$	Significance of the bump	23,618
$A \geq 0.1$ mag	Rejecting low-amplitude variables	
$n_{\mathrm{bump}} = 1$	Rejecting objects with multiple bumps	18,397
	Fit quality:	
$\chi^2_{\mathrm{fit}}/\mathrm{dof} \leq 2.0$	χ^2 for all data	
$\chi^2_{\mathrm{fit},t_{\mathrm{E}}}/\mathrm{dof} \leq 2.0$	χ^2 for $\lvert t - t_0 \rvert < t_{\mathrm{E}}$	
$\sigma(t_{\mathrm{E}})/t_{\mathrm{E}} < 0.5$	Einstein timescale is well measured	
$t_{\mathrm{min}} \leq t_0 \leq t_{\mathrm{max}}$	Event peaked between t_{min} and t_{max}, which are moments of the first and last observation of a given field	
$u_0 \leq 1$	Maximum impact parameter	
$t_{\mathrm{E}} \leq 500\,\mathrm{d}$	Maximum timescale	
$A \geq 0.4$ mag if $t_{\mathrm{E}} \geq 100$ days	Long-timescale events should have high amplitudes	
$I_{\mathrm{s}} \leq 21.0$	Maximum I-band source magnitude	
$F_{\mathrm{b}} > -F_{\mathrm{min}}$	Maximum negative blend flux, corresponding to $I = 20.5$ mag star	460

Mroz, P. et al 2020

What is released

Machine Learning Signatures of Dark Objects

Machine Learning Microlensing

- MicroLIA (Godines, D. et al 2019) a python package (github.com/Professor-G/MicroLIA)
  - Simulation of light curves (of some astrophysical phenomena)
  - Extraction of time series features
  - Machine learning classification between different classes
- Produces 74 **features** from the light curve, and from its derivative (total 148)
- Simulates **galactic** sources suitable for an **OGLE**-inspired study plus a constant class (noise)
  - Cataclysmic Variables (CV)
  - RR Lyrae & Cepheid Variables (VARIABLE)
  - Mira long-period variables (LPV)
  - Point-like microlensing (ML)

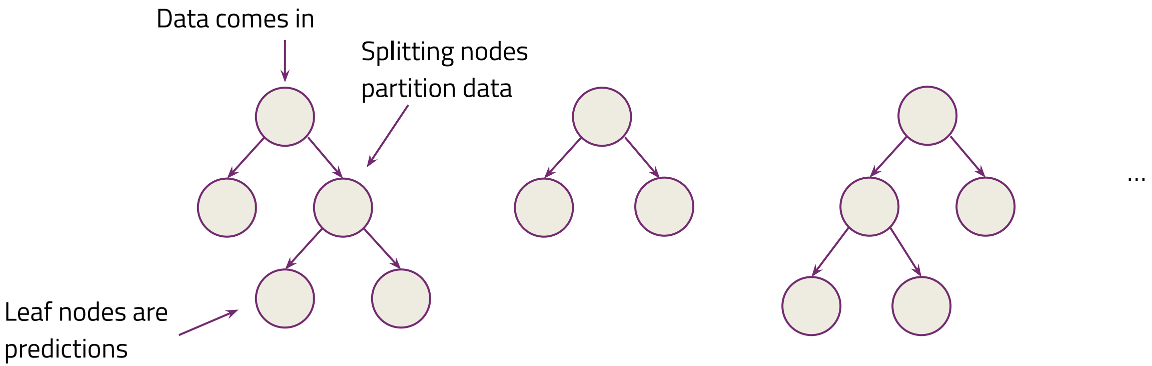
Feature	Description
Above 1‡	Ratio of data points that are above 1 standard deviation from the median.
Above 3	Ratio of data points that are above 3 standard deviations from the median.
Above 5	Ratio of data points that are above 5 standard deviations from the median.
Absolute Energy[†]	The sum over the squared values of the time-series.
Absolute Sum of Changes[†]	The absolute value of the sum over the consecutive changes in the time-series.
Amplitude‡	Difference between the 2^{nd} and 98^{th} percentile of the time-series.
Autocorrelation[†]	Similarity between observations as a function of a time lag between them.
Below 1	Ratio of data points that are below 1 standard deviation from the median.
Below 3	Ratio of data points that are below 3 standard deviation from the median.
Below 5	Ratio of data points that are below 5 standard deviation from the median.
C3[†]	A measure of non-linearity in the time series, introduced by Schreiber and Schmitz (1997).
Check Duplicate[†]	Checks whether any measurements in the time-series repeat at least twice.
Check Max Duplicate[†]	Checks whether the maximum value in the time-series repeats.
Check Min Duplicate[†]	Checks whether the minimum value in the time-series repeats.
Check Max Last Loc[†]	Measures the first location of the maximum value, relative to the length of the time-series.
Check Min Last Loc[†]	Measures the first location of the minimum value, relative to the length of the time-series.
Complexity[†]	Measured by "stretching" the time-series and calculating the length of the resulting line, introduced by Batista et al. (2014).
Con[§]	Number of clusters containing three or more consecutive observations larger than the baseline value plus 3 standard deviations.
Con 2	Number of clusters containing three or more consecutive observations larger than the baseline value plus 2 standard deviations.
Count Above[†]	Number of measurements in the time-series greater than the mean value.
Count Below[†]	Number of measurements in the time-series smaller than the mean value.
First Loc Max[†]	Returns the normalized first location of the maximum value in the time-series.
First Loc Min[†]	Returns the normalized first location of the minimum value in the time-series.
Integrate	Integration of the time-series using the trapezoidal rule.
Kurtosis[‡]	A measure of the peakedness of the lightcurve relative to a normal distribution.
Longest Strike Above[†]	The length of the longest sequence of consecutive measurements in the time-series greater than the mean value.
Longest Strike Below[†]	The length of the longest sequence of consecutive measurements in the time-series smaller than the mean value.
Mean Absolute Change[†]	The mean over the absolute differences between subsequent measurements.
Mean Change[†]	The mean over the differences between subsequent measurements.
Mean Second Derivative[†]	The mean value of a central approximation of the second derivative.
Median Absolute Deviation[‡]	Mean average distance between each measurement and the mean value.
Median Buffer Range[‡]	Ratio of points that are between ± 20% of the amplitude value over the mean.
Median Buffer Range 2	Ratio of points that are between ± 10% of the amplitude value over the mean.
Peak Detection	Calculates the number of peaks in the time-series.
Ratio of Recurring Points[†]	Relative number of time-series values that appear more than once.
Root Mean Squared	The root mean square deviation of the time-series.
Sample Entropy[†]	The sample entropy of the time-series as developed by Richman and Moorman (2000).
Shannon Entropy	Measures the amount of information carried by a signal (Shannon and Weaver, 1949).
Skewness[‡‖]	Measures the asymmetry of the time-series.
STD[‡]	The standard deviation of the time-series.
STD Over Mean[§]	Ratio of standard deviation to mean value.
StetsonJ[‡§]	Variability index first suggested by Stetson (1996) which measures the correlation between each measurement.
StetsonK[‡§]	Index first suggested by Stetson (1996) which serves as a robust kurtosis measure.
StetsonL	Variability index first suggested by Stetson (1996) to distinguish between different types of variation.
Sum Values[†]	Sum over all time-series measurements.
Time Reversal Asymmetry[†]	Measures the asymmetry of a series upon time-reversal (Schreiber and Schmitz, 2000).
von Neumann Ratio[§‖]	The mean square successive difference divided by the sample variance.

Godines, D et al 2019

Machine Learning Signatures of Dark Objects

Machine Learning Microlensing

MicroLIA classification step implements a Random Forest classifier: an ensemble method that uses smaller, weaker, learners: small decision trees to produce a strong learner.
Each tree is trained on a subset of the data and recursively partitions it into each of the classes we want to predict



The final prediction is the average of the predictions of the weaker learners (wisdom of the crowd).

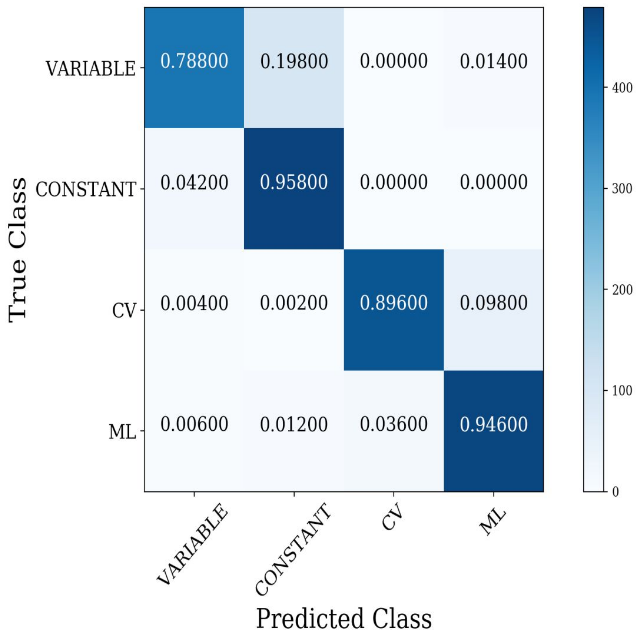
Machine Learning Signatures of Dark Objects
Machine Learning Microlensing

In Godines, D. et al 2019 they found that point-like microlensing events could be isolated from other galactic sources.

This shows a lot of promise! But how do Boson Stars and NFW subhalos (and in general other extended objects) fit in this picture?

- Can we isolate them?
    - Especially from point-like light curves?
- Could they be "polluting" point-like observations?

Godines, D et al 2019



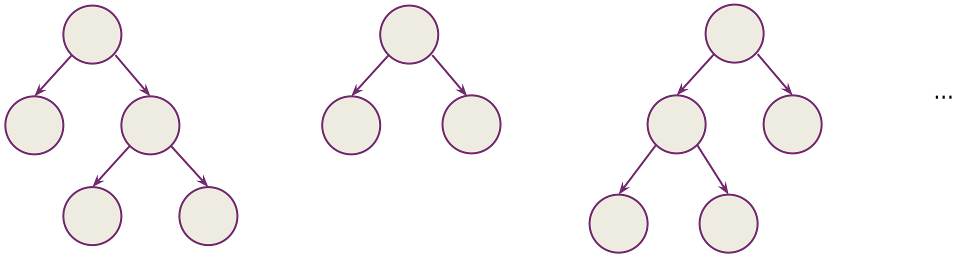
Machine Learning Signatures of Dark Objects
Machine Learning Microlensing

- We extended MicroLIA to simulate BS and NFW light curves
  - Same cadence
  - Same noise model (Gaussian)
- Improved MicroLIA features extraction pipeline
  - Corrected an error with derivative time series implementation
  - Adapted the code for parallel simulation
  - Demanded that $\mu \geq 1.34$ for at least one timestamp
- Produced two datasets each with 100k light curves per class
  - OGLE2 timestamps
    - Adaptive cadence using OGLE2 timestamps of real events
    - Mimics OGLE2 survey sensitivity
  - “Perfect” daily cadence timestamps
    - Ideal case to study the impact of cadence
- Dataset released with the paper doi.org/10.5281/zenodo.10566869

Machine Learning Signatures of Dark Objects
Machine Learning Microlensing

For the classification task we used a Histogram-Based Gradient Boosted Machine, another type of ensemble learners (implemented using `scikit-learn`).

In Random Forests, the weak learners are independent of each other. In Gradient Boosted Machines, the weak learners are trained sequentially on the prediction error of the previous iteration



$F_1(X)$ trained on target y

$F_2(X) = F_1(X) + h_1(X)$
w/ h_1 trained on residual
$y - F_1(X)$

$F_3(X) = F_2(X) + h_2(X)$
w/ h_2 trained on residual
$y - F_2(X)$

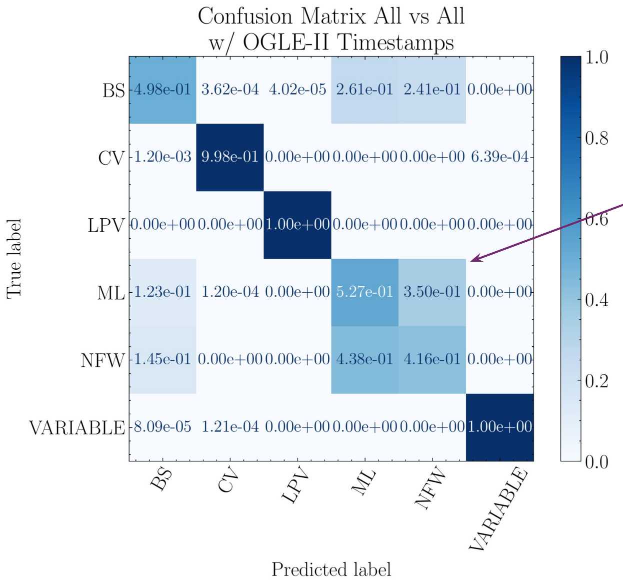
and the final prediction is $F(X)=F_M(X)=F_{M-1}(X)+h_{M-1}(X)$.

Code: gitlab.com/miguel.romao/microlensing-extended-objects-machine-learning

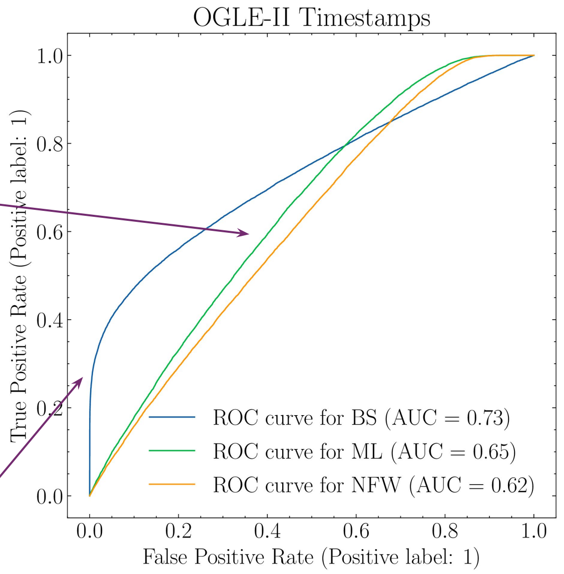
Machine Learning Signatures of Dark Objects
OGLE2 Timestamps Analysis

We find that there is a considerable overlap between microlensing sources.

However, Boson Star light curves appear to be the most distinguishable between them.



NFW and point-like
ML are very mixed



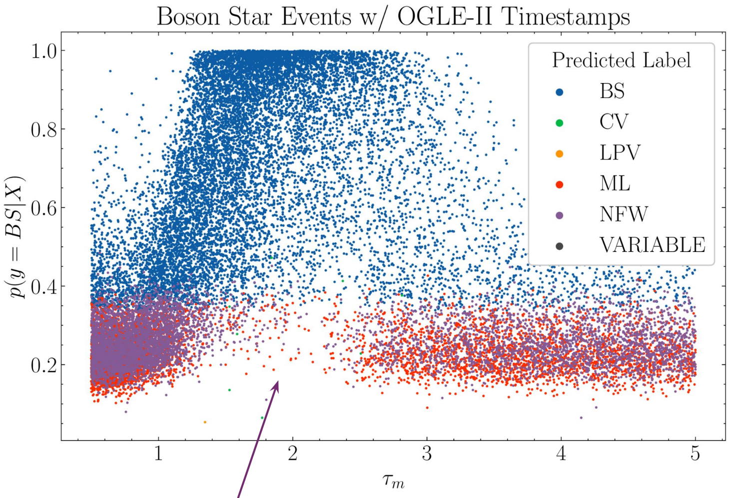
High purity of BS for
a high cut on the
classifier output

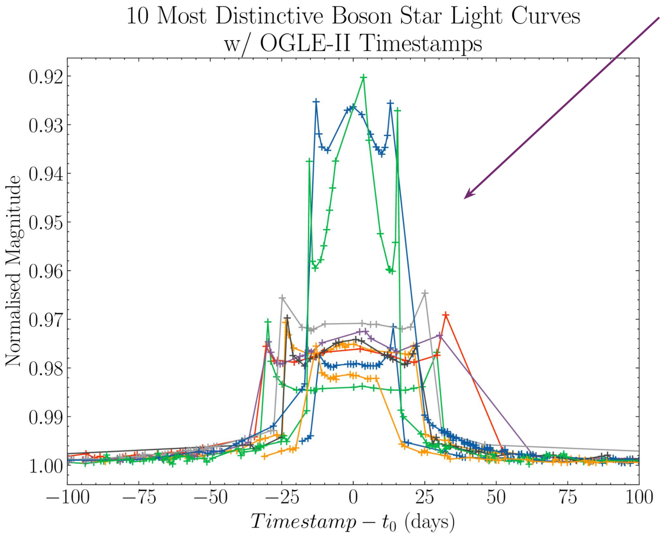
Machine Learning Signatures of Dark Objects

OGLE2 Timestamps Analysis

Indeed, we are able to isolate Boson Stars!

The 10 more confidently classified BS events exhibit the caustics very clearly!





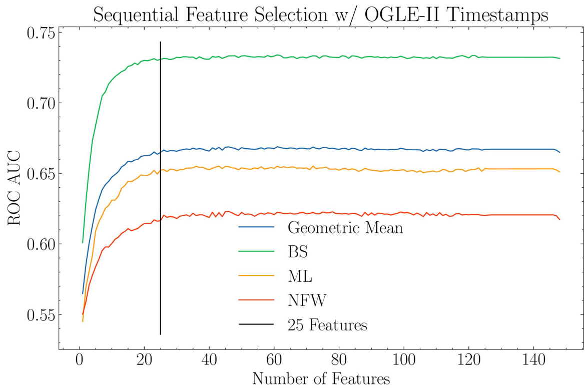
Sweet spot τ_m~2 exactly where the caustics emerge

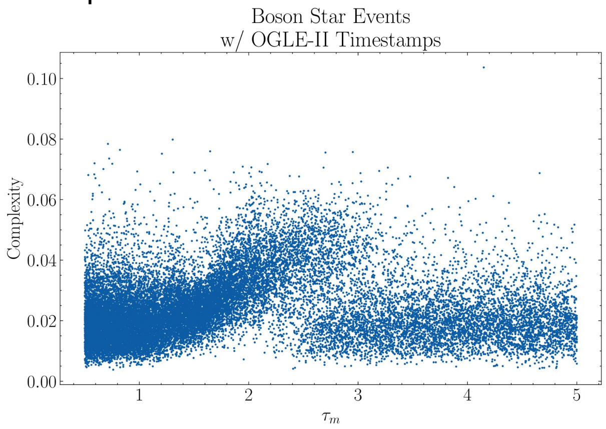
Machine Learning Signatures of Dark Objects
OGLE2 Timestamps Analysis

The input to our classifier were 148 features, so which ones are relevant for the task?

Implemented a Backward Sequential Feature Selection loop: removes a feature each step, keeping the best ones. Found that around **25 features** are important to differentiate between microlensing classes!

For Boson Stars, **complexity** was found to be the most important.

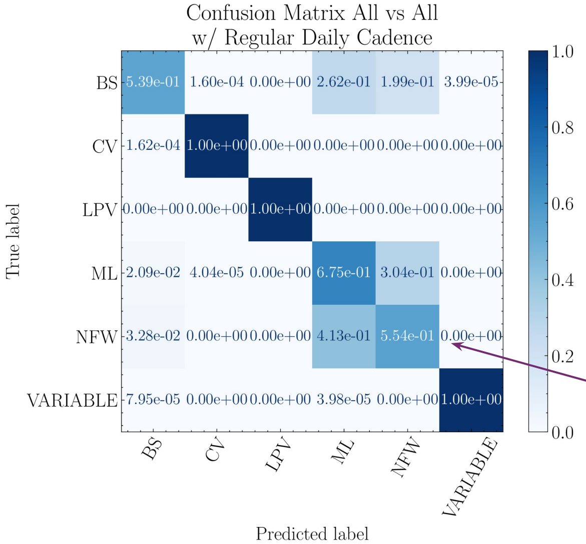




Complexity is the standard deviation of the time series self-difference

Machine Learning Signatures of Dark Objects
Regular Cadence Timestamps Analysis

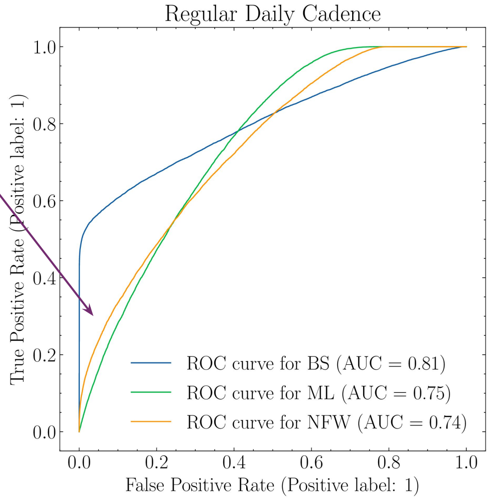
OGLE2 has irregular and low cadence. How much would our results improve if we could have regular daily cadence?



	BS	CV	LPV	ML	NFW	VARIABLE
BS	5.39e-01	1.60e-04	0.00e+00	2.62e-01	1.99e-01	3.99e-05
CV	1.62e-04	1.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00
LPV	0.00e+00	0.00e+00	1.00e+00	0.00e+00	0.00e+00	0.00e+00
ML	2.09e-02	4.04e-05	0.00e+00	6.75e-01	3.04e-01	0.00e+00
NFW	3.28e-02	0.00e+00	0.00e+00	4.13e-01	5.54e-01	0.00e+00
VARIABLE	7.95e-05	0.00e+00	0.00e+00	3.98e-05	0.00e+00	1.00e+00

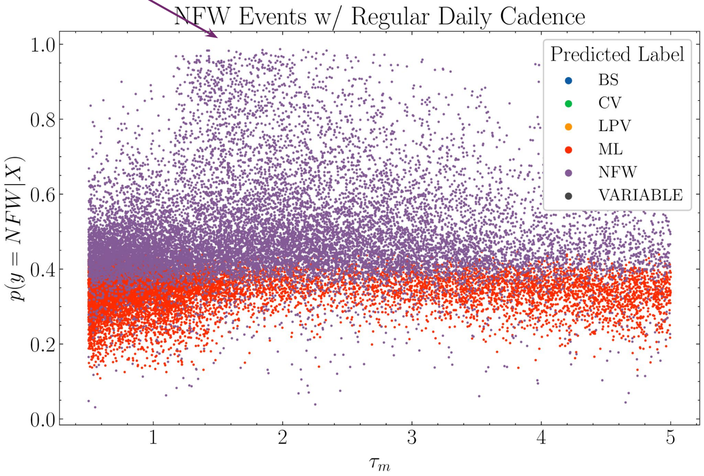
NFW seem better isolated!

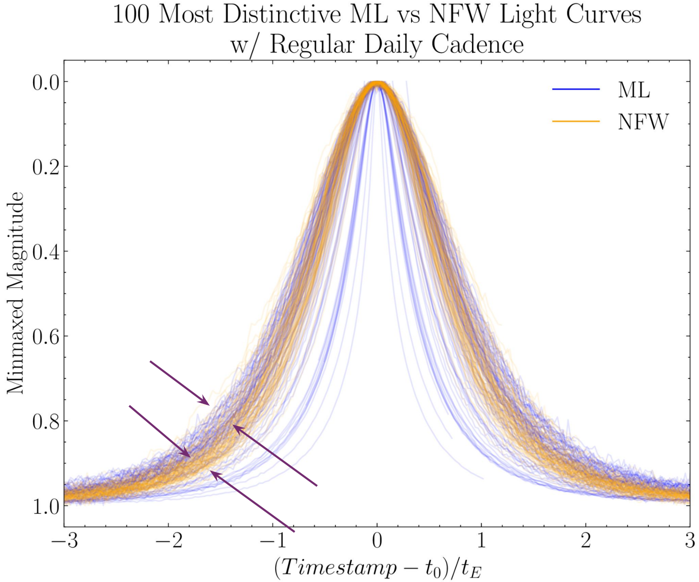
Easier to isolate NFW at low false positive rate cuts!



Machine Learning Signatures of Dark Objects
Regular Cadence Timestamps Analysis

Evidence at the possibility of isolating NFW events with high confidence





Harder to interpret what is happening, but it seems that our classifier is picking up “narrower” peaks

Current Work

Ongoing Work
Motivation

Our first study on machine learning microlensing of extended objects was very exciting! However, a few things stand out:

- Our dataset does not include possible sources of misclassification, (e.g. binary lenses)
- Our simulation made use of a simplistic noise model
- The methodology is detached from what surveys do
  - Surveys implement cuts based on point-like microlensing light curve profiles
  - No machine learning is applied to the light curves themselves, which are a by-product of analyses
- How do we take our message across?
  - We need to bring our extended object light curve models to the astronomer community
  - We need to expand our analysis to include more realistic noise models and simulations
  - Look for ongoing community efforts in astrophysics

In a sense: we need to reach out to astronomers, especially those associated with surveys!

Ongoing Work
Reaching out to Astronomers

We (Djuna and I) have teamed up with MicroLIA's main author Daniel Godines, and our first stop is the ELAsTiCC dataset:

- **E**xtended **L**SST **As**tronomical **Ti**me Series **C**lassification **C**hallenge
    - LSST: Legacy Survey of Space and Time
        - To be conducted by Vera C. Rubin Observatory
    - Inherited parts of PLAsTiCC (P for photometric), Kessler, R et all 2019
        - And its Kaggle challenge Hlozek, R et all 2023
- Multiple sources, goth galactic and extragalactic
    - However, a clear focus on supernovae as ELAsTiCC is hosted by the DESC: Dark Energy Science Collaboration
- Science purposed
  ELAsTiCC presents the first simulation of LSST alerts, with millions of synthetic transient light curves and host galaxies. The data is being used to test broker alert systems and classifiers, and develop the infrastructure for LSST's Dark Energy Science Collaboration Time-Domain needs.

Ongoing Work
Reaching out to Astronomers

Dataset available online: portal.nersc.gov/cfs/lsst/DESC_TD_PUBLIC/ELASTICC/

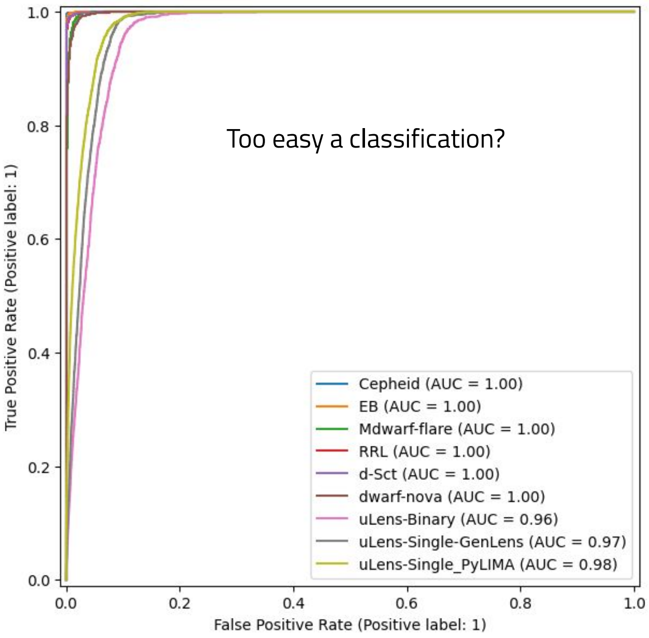
- Includes O(2M) labeled simulated light curves
  - Of which O(200k) galactic sources
    - Eclipsing Binaries (94934)
    - δ-Scuti (29499)
    - RR Lyrae (20047)
    - Cepheid Variables (19672)
    - Dwarf Nova (11464)
    - Microlensing Point-like Lens, PyLIMA (10635)
    - Microlensing Point-like Lens, GenLens (9371
    - Microlensing Binary Lens (5123)
    - M-dwarf flare (2655)

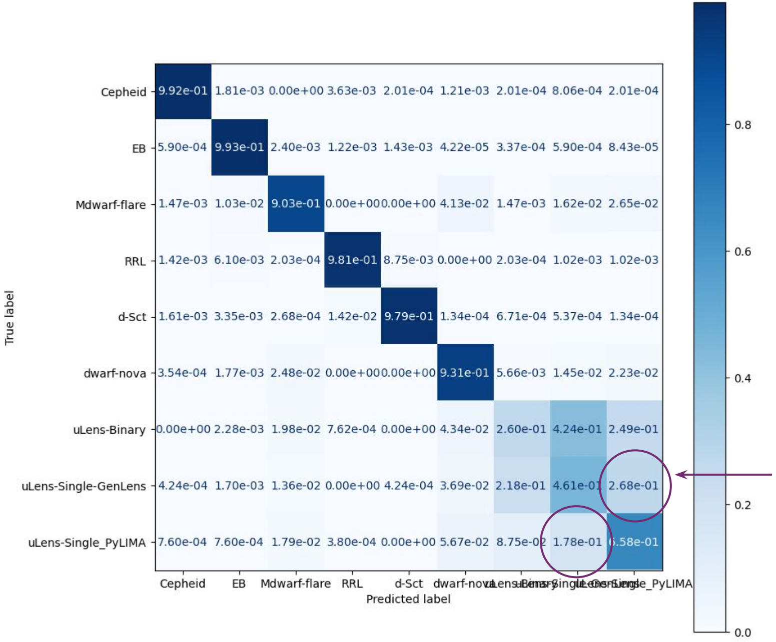
Two different simulations of point-like microlensing

We are very interested in new classes like this

Ongoing Work
Reaching out to Astronomers

Preliminary classification study on ELAsTiCC

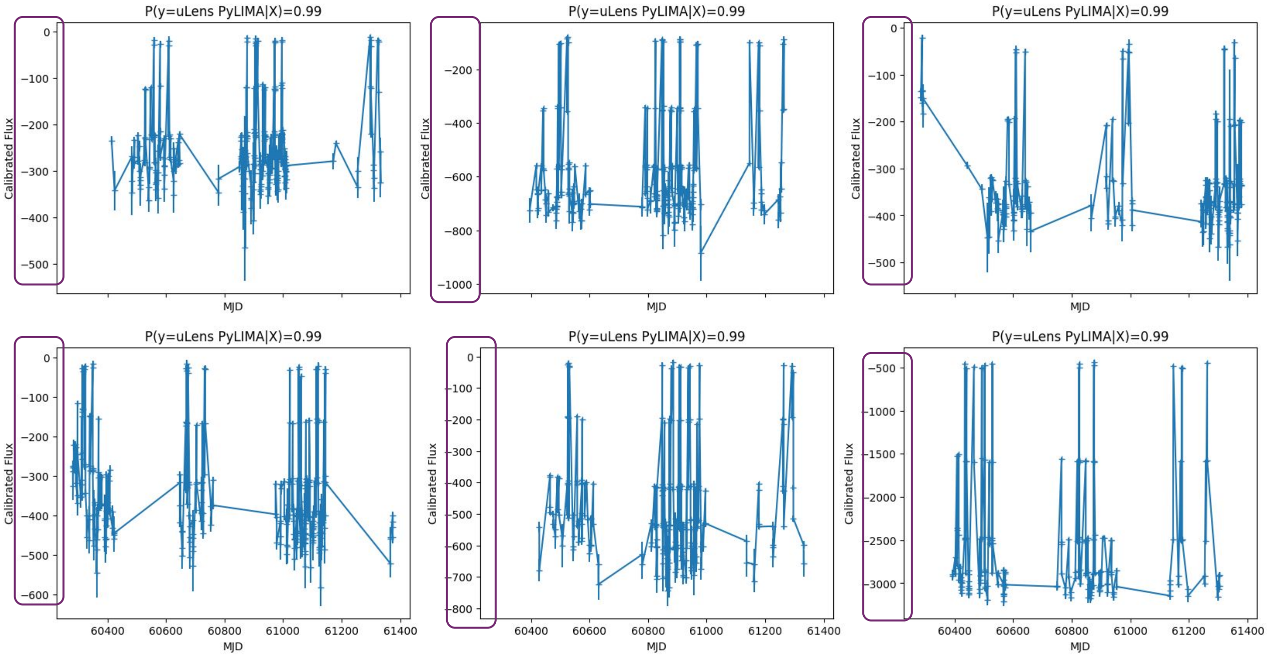




This is odd

Ongoing Work
Not all is well

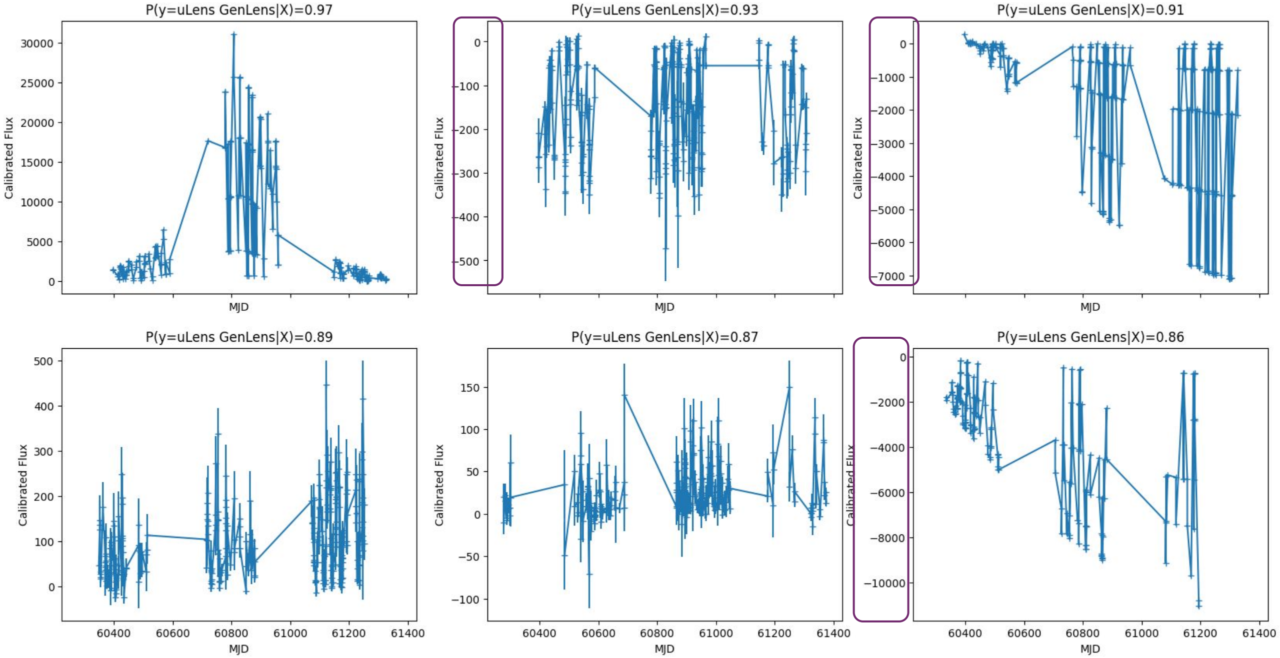
The classifier is picking up simulation differences, not Physics…



Ongoing Work
Not all is well

The classifier is picking up simulation differences, not Physics…

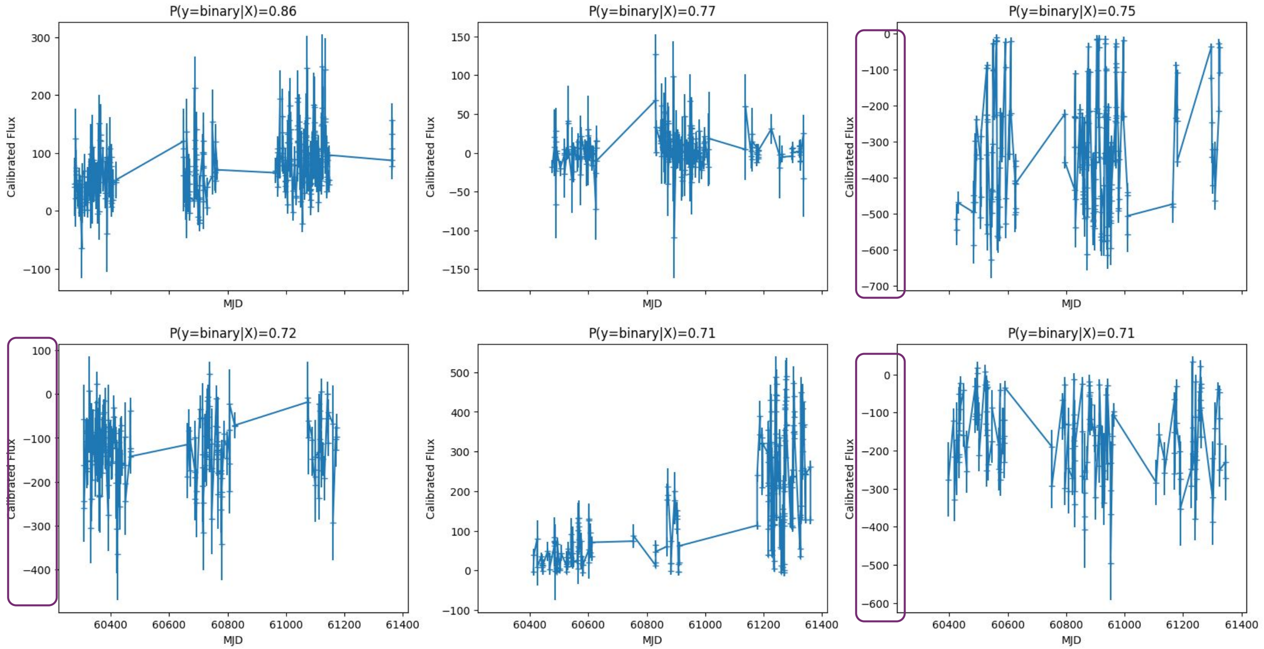
Maybe this is a microlensing light curve?



But one needs a lot more imagination for these

Ongoing Work
Not all is well

The classifier is picking up simulation differences, not Physics…



Likewise here…

Ongoing Work
Going the extra mile

- In fact, things get **worse**... using the `rubin-sim` (github.com/lsst/rubin_sim) package (since ELAsTiCC is streaming to brokers, **no simulation code is available**), which **simulates Vera Rubin Observatory conditions and noise model**, we have simulated point-like microlensing events and **inject them into ELAsTiCC dataset**
  - The classifier was capable to **completely separate** our simulation of point-like microlensing light curves and those present in ELAsTiCC...
- Unfortunately, this means that we cannot “bring your own light curve models” and produce new ELAsTiCC-compatible light curves and perform an analysis with that data...
- So what now?
  - While we cannot inject our extended object light curves into ELAsTiCC, we can produce a dataset of microlensing light curves adapted to Vera Rubin Observatory
  - Stay tuned...

Conclusions and Outlook

Conclusions
And outlook

- Dark matter composition might include extended dark objects, such as NFW subhalos and Boson Stars
- These objects produce **microlensing events**, which can be detected by surveys
- The extended nature of these objects endows their light curves with **unique signatures**, such as caustics
- Using **machine learning** on the time-series features of their light curves, these **objects could, in fact, be discovered** using microlensing surveys
    - Boson Stars exhibit a distinct profile
    - NFW subhalos might require higher (at least more regular) cadence
    - Interestingly, our results did not extend the feature set: in principle better sensitivity could be achieved with new dedicated features

Conclusions
And outlook

- Moving forward we need to close the gap between our phenomenological study and survey analyses
  - Approaching what the community is doing (especially around LSST)
  - Deploying more realistic survey noise models
  - Broadening the classification analysis to include new classes of light curves
- Surveys seem to have too strict a selection criterion
  - Are we missing out on (more) exotic objects?
  - How feasible would be a "model independent" search?
- Finally, there is the exciting prospect that NFW light curves could be isolated
  - Specific time series features that capture some unique signature?
  - Go beyond the features, what could we do with the light curve itself?

Thank you!

n+1

n+1

Backups

Simulation details:

Gaussian (in fact, Poissonian) noise

$\tau_m \sim U(0.5,5)$

$t_E \sim N(\mu=30,\sigma=10)$

$$u_0 \sim \begin{cases} \mathcal{U}(0,1) & \text{for ML} \\ \mathcal{U}(0,1.5) & \text{for BS} \\ \mathcal{U}(0,1.1) & \text{for NFW} \end{cases}$$

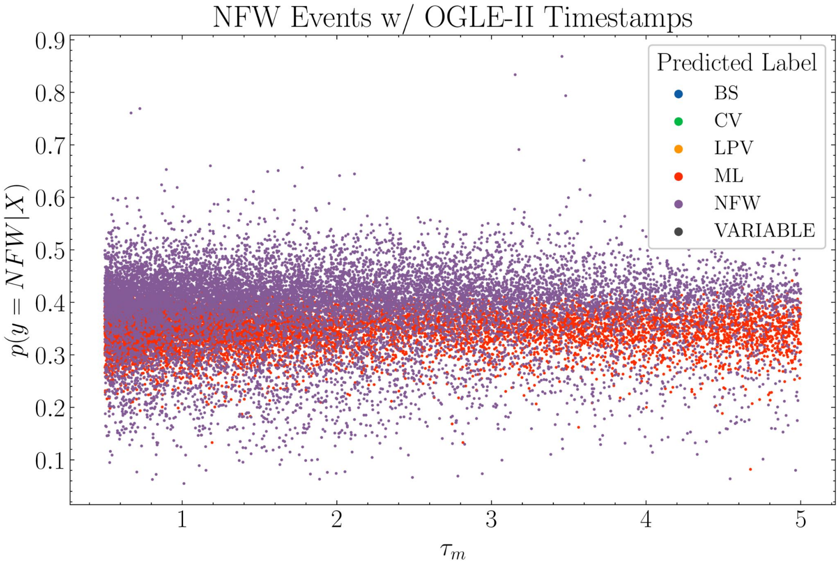
Cuts:

$\mu \geq 1.34$

And MicroLIA criteria: magnitude greater than 3σ at least for 1/3 timestamps, etc

These are minimal cuts, no fits to microlensing curves are performed.

n+1
Backups

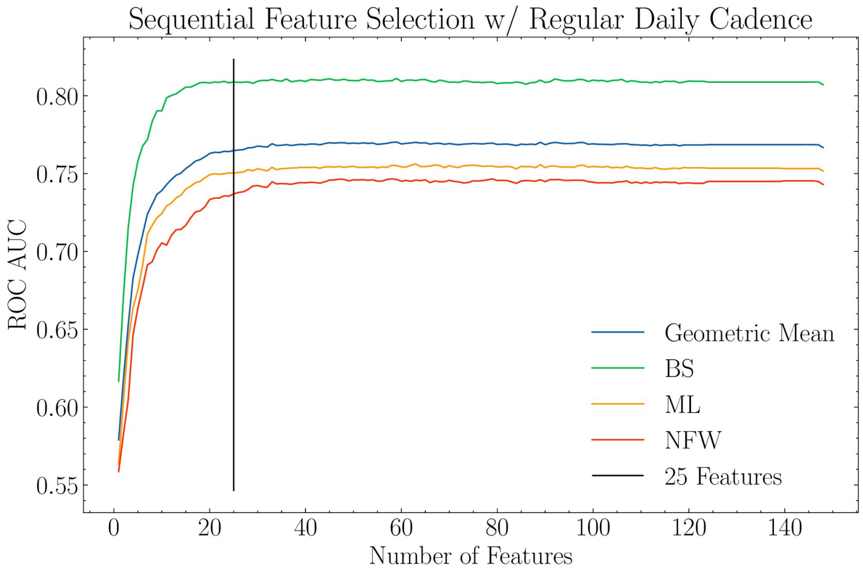


NFW Events w/ OGLE-II Timestamps
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
$p(y = NFW|X)$
1
2
3
4
5
τ_m
Predicted Label
BS
CV
LPV
ML
NFW
VARIABLE

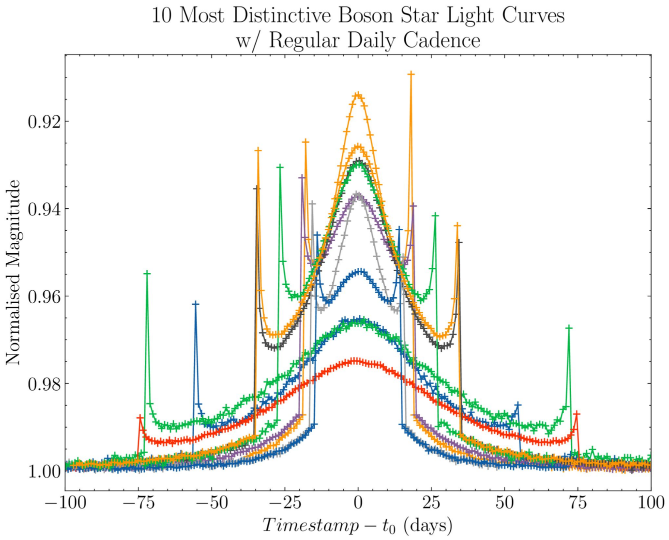
n+1
Backups

Feature Rank	OGLE-II	Daily Regular Cadence
1	complexity	stetsonJ (derivative)
2	median buffer range	half mag amplitude ratio
3	ratio recurring points	FluxPercentileRatioMid80
4	amplitude	median buffer range
5	medianAbsDev	count below
6	FluxPercentileRatioMid80	medianAbsDev
7	count below	check max last loc
8	cusum	longest strike below
9	shapiro wilk	complexity
10	mean change	PercentDifferenceFluxPercentile
11	longest strike above (derivative)	longest strike above
12	mean second derivative	shapiro wilk (derivative)
13	number of crossings	FluxPercentileRatioMid65
14	LinearTrend	number of crossings
15	longest strike below (derivative)	number cwt peaks
16	sample entropy	number cwt peaks (derivative)
17	FluxPercentileRatioMid65	time reversal asymmetry (derivative)
18	mean change(derivative)	FluxPercentileRatioMid35
19	longest strike above	peak detection(derivative)
20	mean n abs max (derivative)	mean n abs max (derivative)
21	FluxPercentileRatioMid20	longest strike above (derivative)
22	quantile (derivative)	check min last loc
23	check min last loc	quantile (derivative)
24	Gskew	sample entropy (derivative)
25	check max last loc (derivative)	below3 (derivative)

n+1
Backups



Sequential Feature Selection w/ Regular Daily Cadence
ROC AUC
0.80
0.75
0.70
0.65
0.60
0.55
0
20
40
60
80
100
120
140
Number of Features
Geometric Mean
BS
ML
NFW
25 Features



10 Most Distinctive Boson Star Light Curves
w/ Regular Daily Cadence
Normalised Magnitude
0.92
0.94
0.96
0.98
1.00
−100
−75
−50
−25
0
25
50
75
100
$Timestamp - t_0$ (days)