UKIRT Microlensing Survey as a Pathfinder for WFIRST

Geoffrey Bryden
Jet Propulsion Laboratory

UKIRT microlensing team:
Yossi Shvartzvald, Savannah Jacklin,
Sebastiano Calchi Novati, Calen Henderson, Scott Gaudi,
Matthew Penny, David Nataf, Chas Beichman,
Kiri Wagstaff, Selina Chu

JPL

OHIO STATE

VANDERBILT

NExScI
NASA Exoplanet Science Institute

A Near-IR Survey with UKIRT

I-band microlensing surveys

OGLE

MOA

KMTNet

near-IR microlensing surveys

UKIRT

3.8m NIR telescope @ Mauna Kea

WFIRST

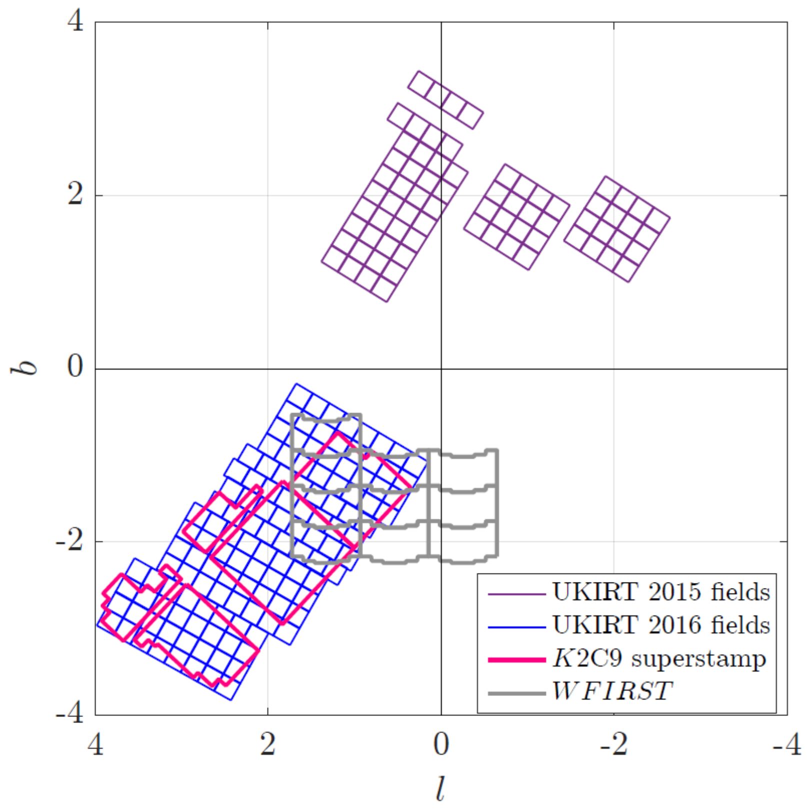
UKIRT 2015-2016 microlensing surveys

2015 survey – *Spitzer*:

- Area: 3.4 deg^2
- Duration: 39 nights
- Cadence: 5 epochs/night
- Total epochs per field: ~145
- Filter: *H*

2016 survey – *K2C9*:

- Area: 6.0 deg^2
- Duration: 91 nights
- Cadence: 2-3 epochs/night
- Total epochs per field: ~160
- Filter: *H*

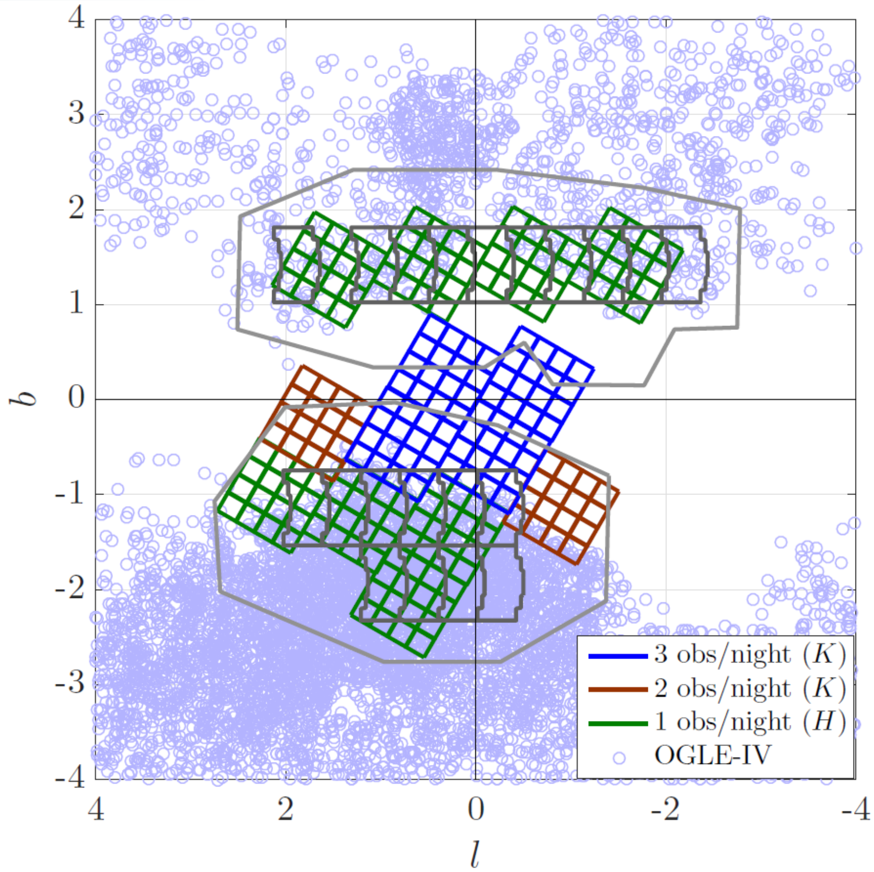


Shvartzvald et al. 2017

UKIRT 2017-2019 microlensing survey

2017-2019 survey – WFIRST:

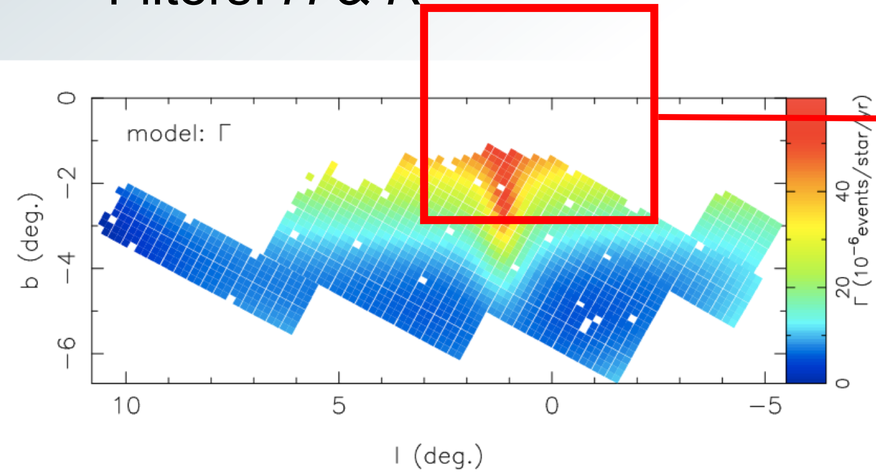
- Area: 10.5 deg^2
- ~20 million lightcurves
- Duration: ~4 months / season
- Cadence: 1-3 epochs / night
- Filters: *H* & *K*



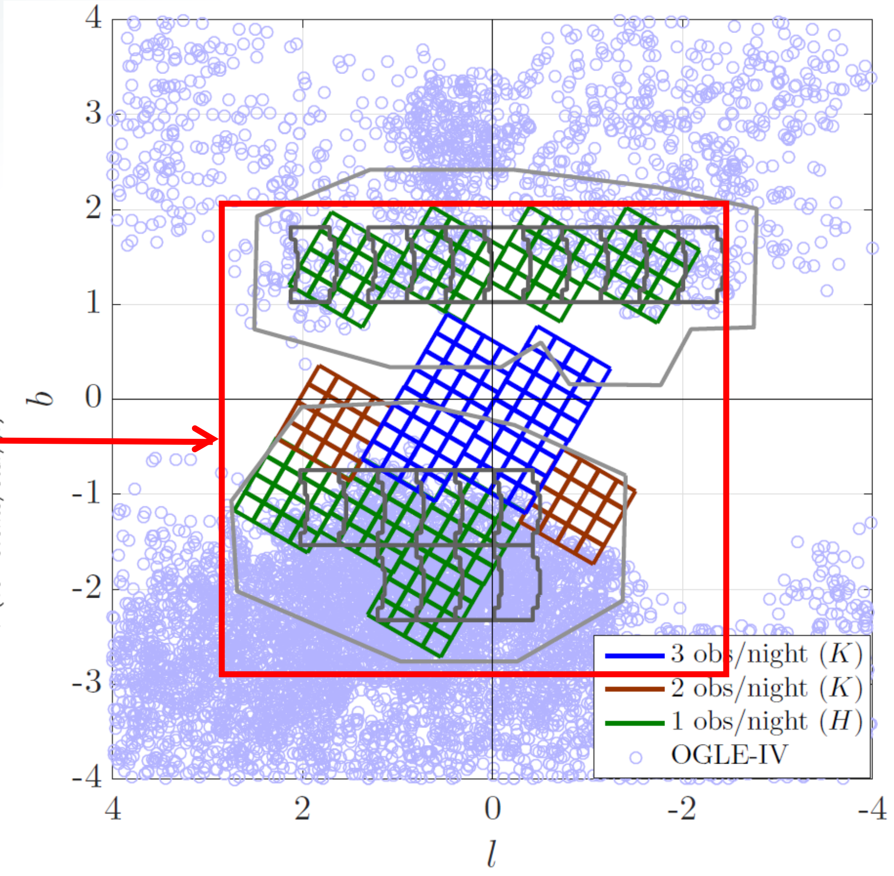
UKIRT 2017-2019 microlensing survey

2017-2019 survey – WFIRST:

- Area: 10.5 deg^2
- ~20 million lightcurves
- Duration: ~4 months / season
- Cadence: 1-3 epochs / night
- Filters: *H* & *K*



Sumi & Penny 2016



By-Eye Results

Unfortunately we cannot completely escape from by-eye analysis. The computer has to learn from the labeled examples we provide.

- We start with an event finder similar to KMT (Kim et al. 2018), based on a 2-D (t_0, t_{eff}) grid search
- **Manual UKIRT Lightcurve Evaluator (MULE)** – a python-based GUI to assist with the by-eye vetting of lightcurves

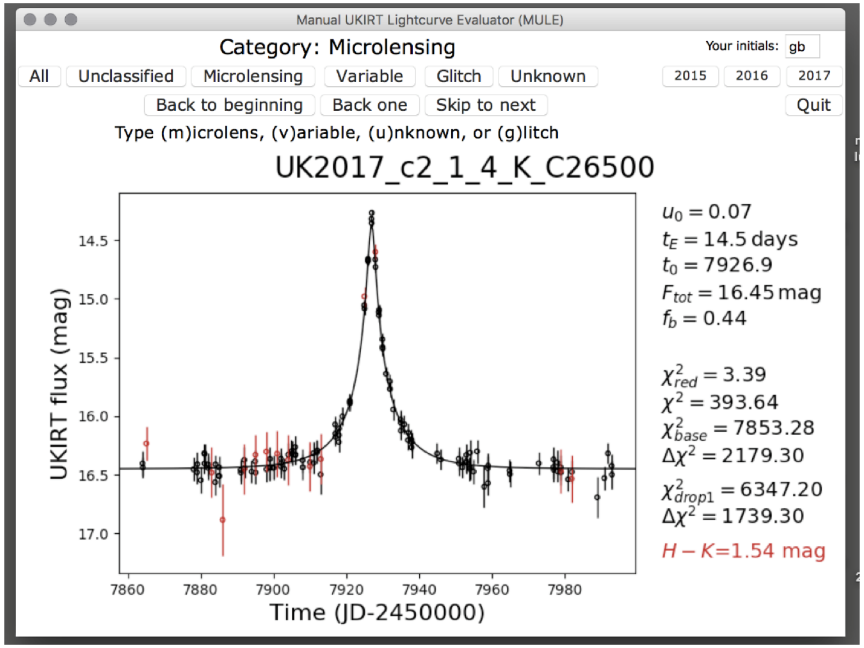
We look at everything with $\Delta\chi^2 > 500$.

2015: 23 events (out of 562)

2016: 65 events (out of 844)

2017: 177 events (out of 2852)

2018: 83 events (out of 843)



Category: Microlensing

Your initials: gb

Type (m)icrolens, (v)ariable, (u)nknown, or (g)litch

$u_0 = 0.07$

$t_E = 14.5$ days

$t_0 = 7926.9$

$F_{tot} = 16.45$ mag

$f_b = 0.44$

$\chi^2_{red} = 3.39$

$\chi^2 = 393.64$

$\chi^2_{base} = 7853.28$

$\Delta\chi^2 = 2179.30$

$\chi^2_{drop1} = 6347.20$

$\Delta\chi^2 = 1739.30$

$H - K$=1.54 mag

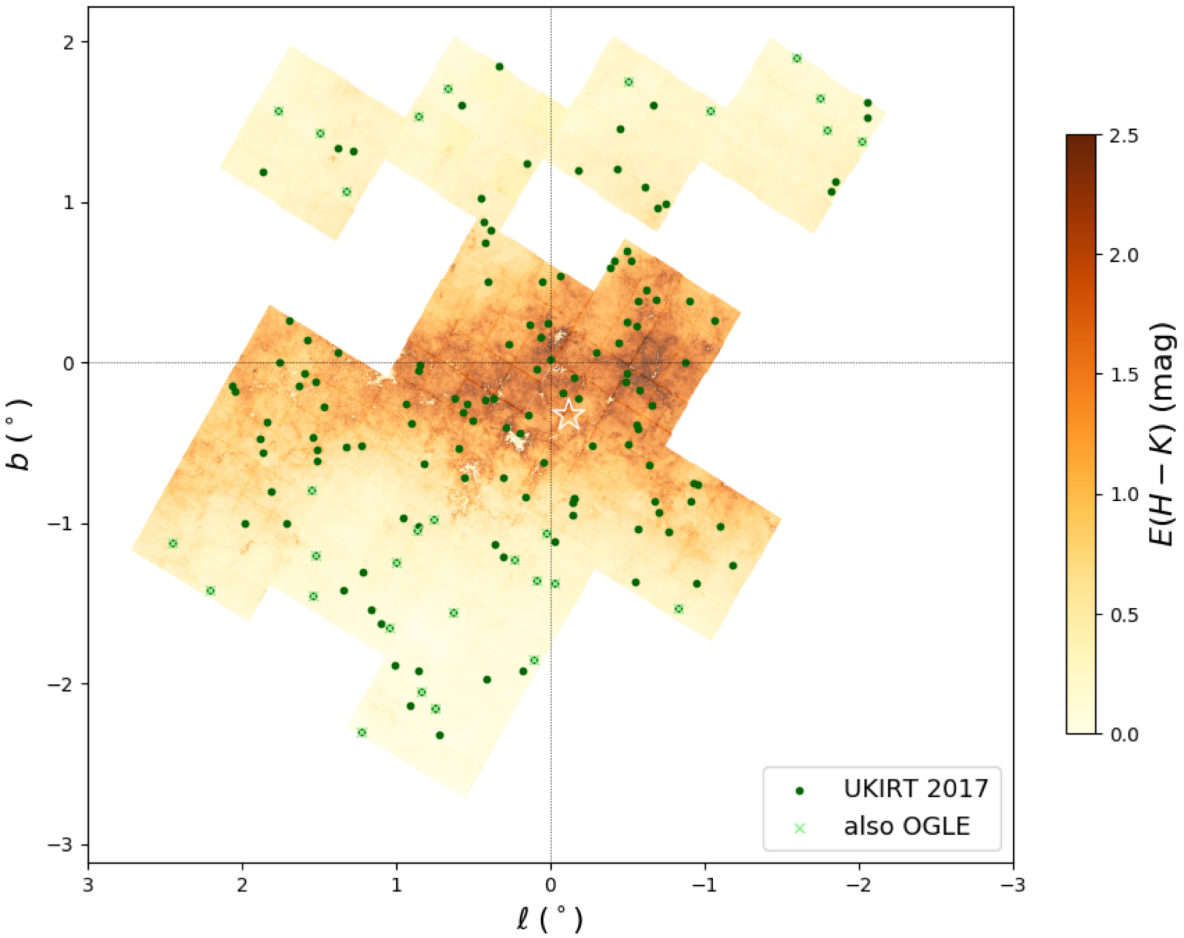
UKIRT microlensing events

2015: 23 events

2016: 65 events

2017: 177 events

2018: 83 events

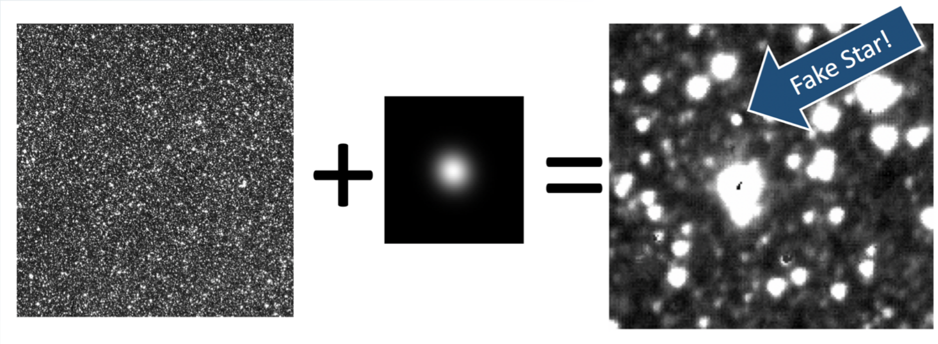


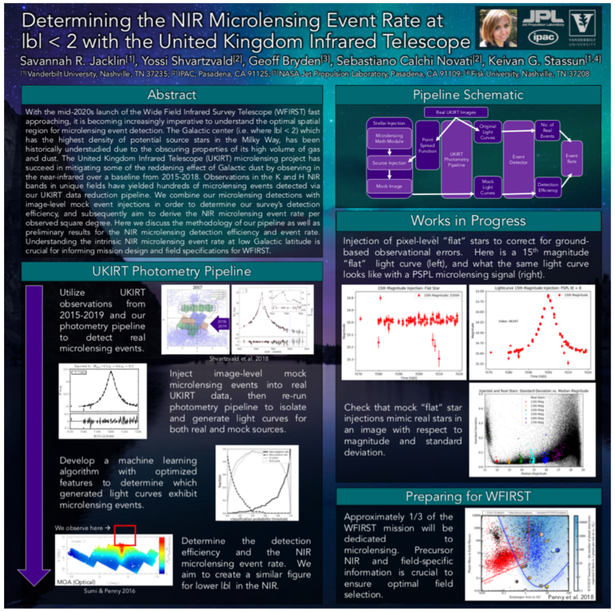
Detection efficiency

There is no way to calculate detection rates from a by-eye selection. Detection efficiency must be calculated based on some well-defined selection criteria – either machine learning or strict cuts.

Detection efficiency is calculated via simulated event injection/recovery.

- Inject events using PSF templates from PSFEx
- Run through full pipeline, including machine learning event detection
- Repeat for many stars, covering parameter space





Determining the NIR Microlensing Event Rate at |b| < 2 with the United Kingdom Infrared Telescope

Savannah R. Jacklin[1], Yossi Shvartzvald[2], Geoff Bryden[3], Sebastiano Calchi Novati[2], Keivan G. Stassun[1,4]
[1] Vanderbilt University, Nashville, TN 37235; [2] IPAC, Pasadena, CA 91125; [3] NASA Jet Propulsion Laboratory, Pasadena, CA 91109; [4] Fisk University, Nashville, TN 37208

Abstract

With the mid-2020s launch of the Wide Field Infrared Survey Telescope (WFIRST) fast approaching, it is becoming increasingly imperative to understand the optimal spatial region for microlensing event detection. The Galactic center (i.e. where |b| < 2) which has the highest density of potential source stars in the Milky Way, has been historically understudied due to the obscuring properties of its high volume of gas and dust. The United Kingdom Infrared Telescope (UKIRT) microlensing project has succeed in mitigating some of the reddening effect of Galactic dust by observing in the near-infrared over a baseline from 2015-2018. Observations in the K and H NIR bands in unique fields have yielded hundreds of microlensing events detected via our UKIRT data reduction pipeline. We combine our microlensing detections with image-level mock event injections in order to determine our survey's detection efficiency, and subsequently aim to derive the NIR microlensing event rate per observed square degree. Here we discuss the methodology of our pipeline as well as preliminary results for the NIR microlensing detection efficiency and event rate. Understanding the intrinsic NIR microlensing event rate at low Galactic latitude is crucial for informing mission design and field specifications for WFIRST.

UKIRT Photometry Pipeline

Utilize UKIRT observations from 2015-2019 and our photometry pipeline to detect real microlensing events.

Shvartzvald et al. 2018

Inject image-level mock microlensing events into real UKIRT data, then re-run photometry pipeline to isolate and generate light curves for both real and mock sources.

Develop a machine learning algorithm with optimized features to determine which generated light curves exhibit microlensing events.

We observe here →

MOA (Optical)

Sumi & Penny 2016

Determine the detection efficiency and the NIR microlensing event rate. We aim to create a similar figure for lower |b| in the NIR.

Pipeline Schematic

Works in Progress

Injection of pixel-level "flat" stars to correct for ground-based observational errors. Here is a 15th magnitude "flat" light curve (left), and what the same light curve looks like with a PSPL microlensing signal (right).

Check that mock "flat" star injections mimic real stars in an image with respect to magnitude and standard deviation.

Preparing for WFIRST

Approximately 1/3 of the WFIRST mission will be dedicated to microlensing. Precursor NIR and field-specific information is crucial to ensure optimal field selection.

Penny et al. 2018

Savannah Jacklin
PhD student
Vanderbilt

Machine Learning for Event Detection

Machine learning has the potential to improve

- efficiency in detecting events
- consistency in detecting events

This is particularly true for larger datasets (WFIRST).

Meanwhile we are using UKIRT analysis as a pathfinder.

Goal 1: Save time in detecting microlensing events

Goal 2: Enable robust, efficient detection statistics

(cf. Wyrzykowski+ 2015: machine learning for OGLE data)

Machine-Learning Event Detection

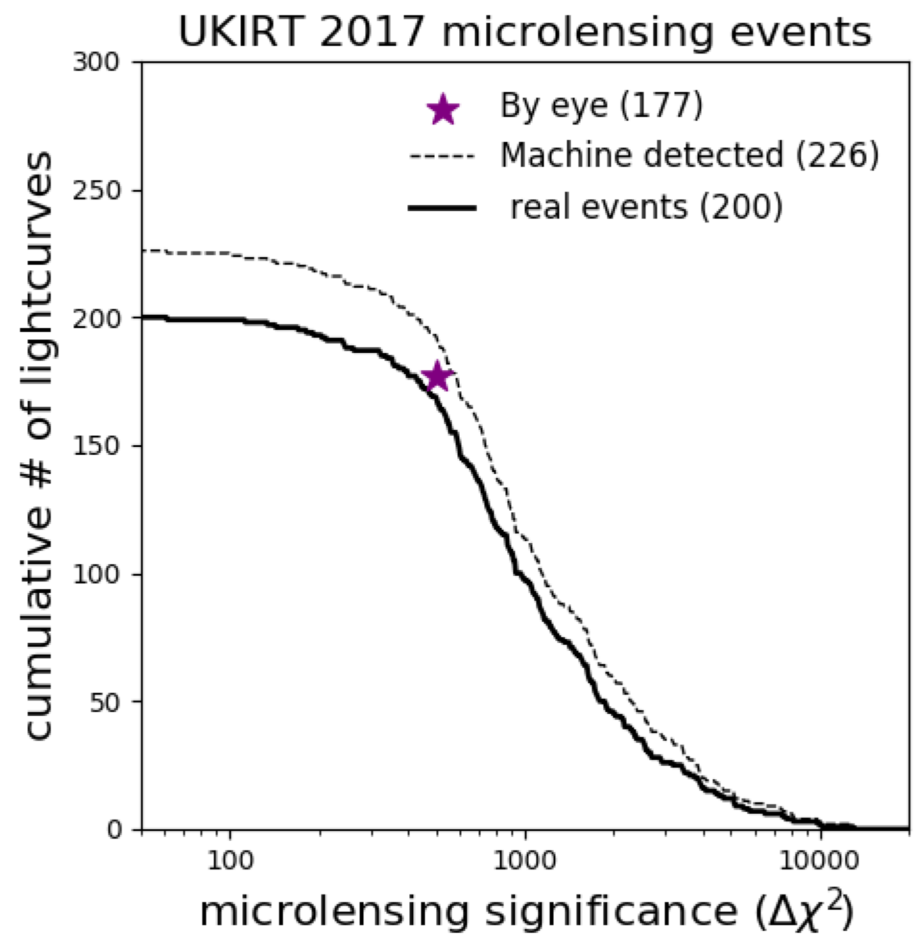
How well do we do, compared to by-eye selection?

Only a modest increase in the number of events detected (177 → 200).

A handful of the original detections were missed by the machine learning scheme.

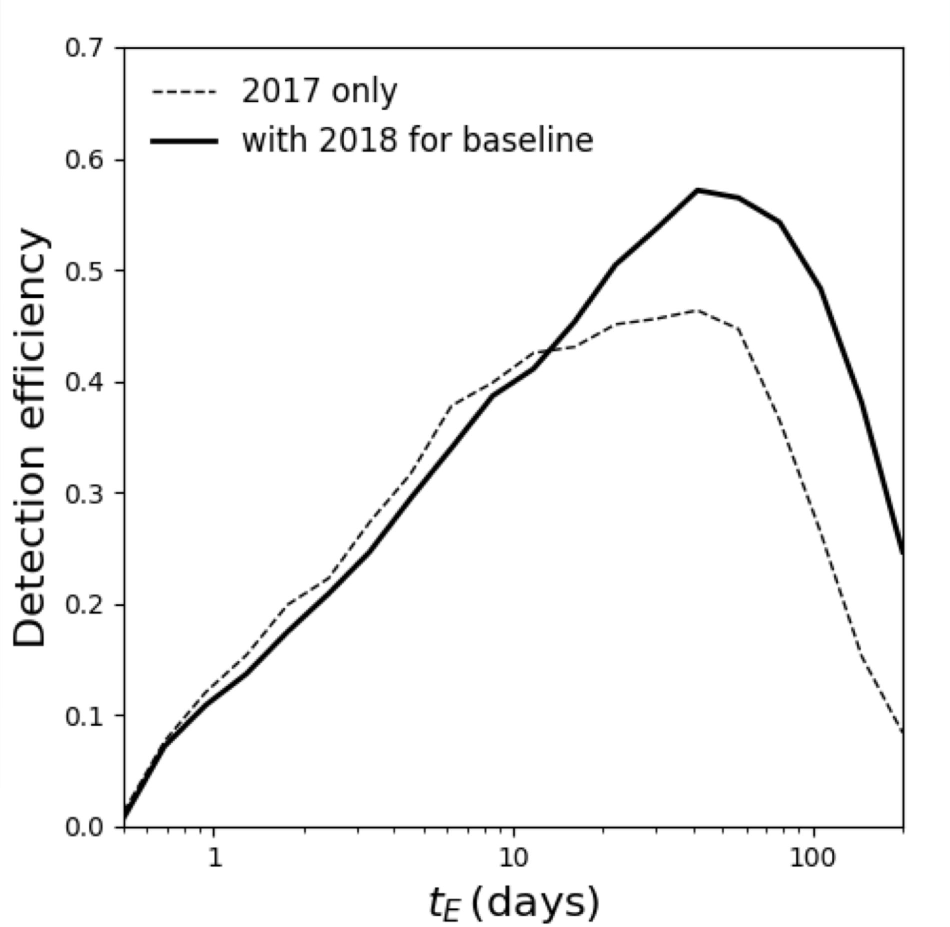
There were 26 false detections (variable stars that look similar to microlensing events).

(cf. OGLE; Wyrzykowski+ 2015)



Machine-Learning Detection Efficiency

Injections of events into images are in progress.
Meanwhile we are injecting events directly into light curves.



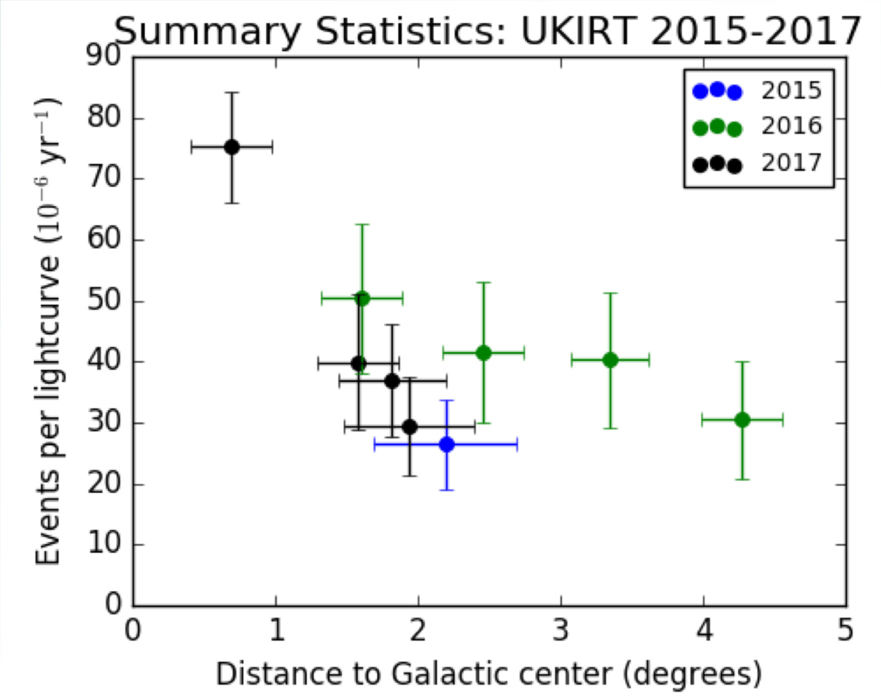
→ up to 60% detection efficiency

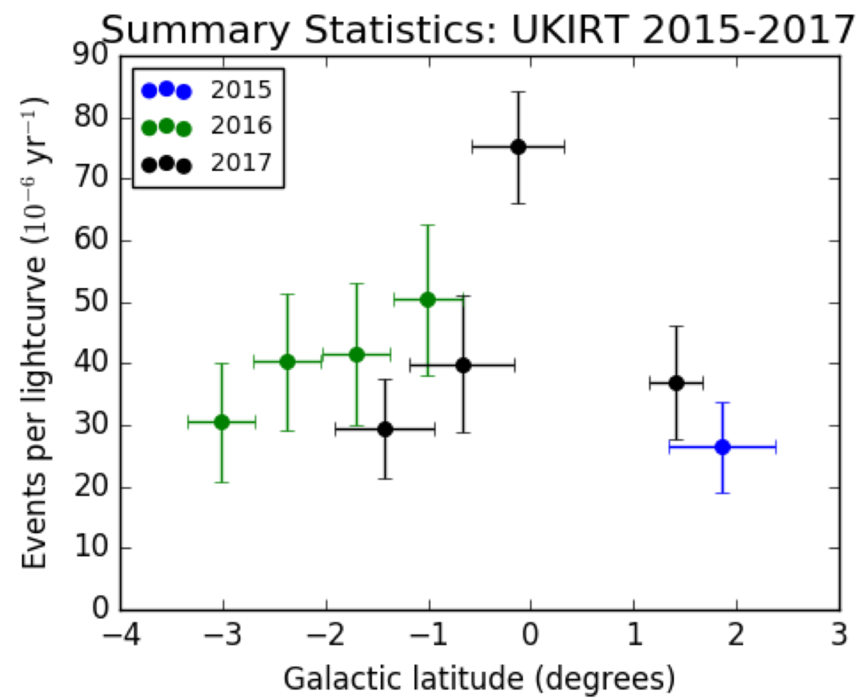
This is very high,
particularly when considering the limitations of the UKIRT dataset (just 2 short seasons).

Near-IR event rate

Preliminary results:

1. High event rate in the central fields
2. No excess of events in the northern bulge





Additional Science

2015:

- A massive remnant in wide binary:

  OGLE-2015-1285 (Shvartzvald et al. 2015)

2016:

- Planets:

  MOA-2016-227 (Koshimoto et al. 2017)

  OGLE-2016-0163 (Han et al. 2017)

  OGLE-2016-1190 (Ryu et al. submitted)

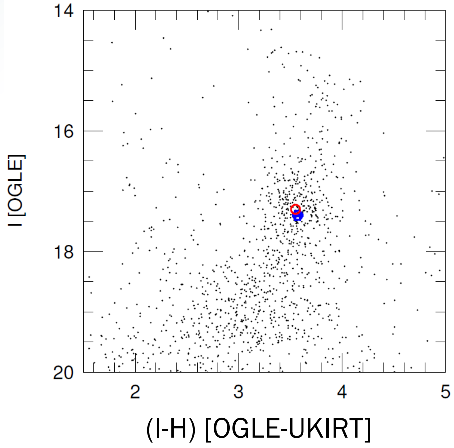
  OGLE-2016-0241 (Poleski et al. in prep.)

2017:

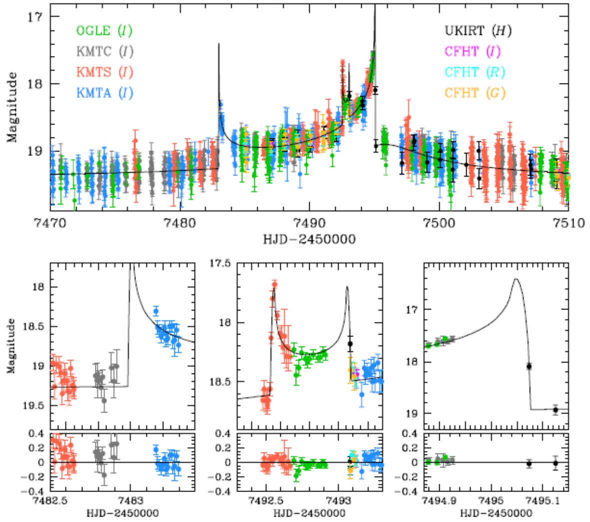
- Planet:

  OGLE-2017-0173 (Hwang et al. 2017)

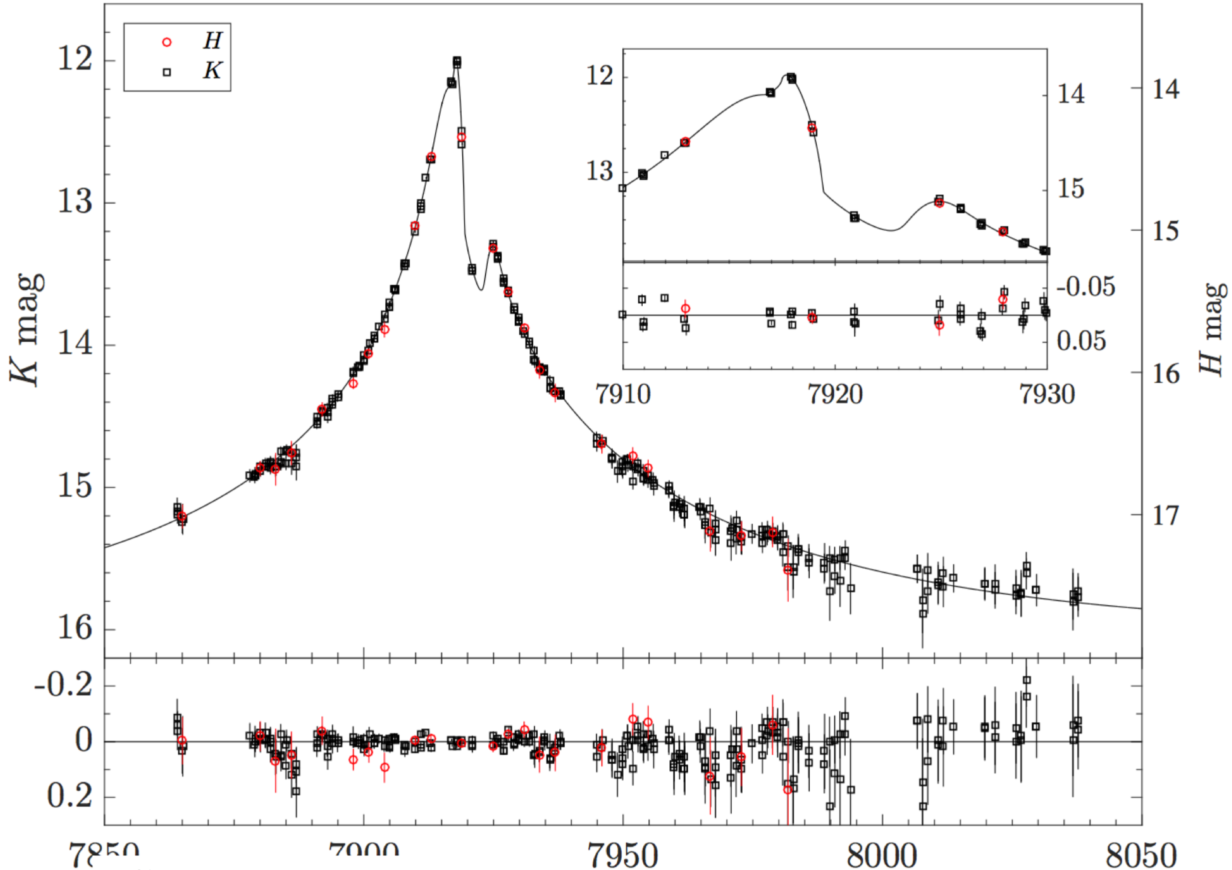
OGLE-15-1285

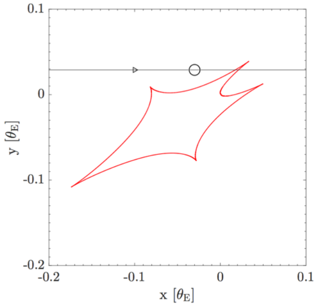


OGLE-16-0163



UKIRT-2017-BLG-001b (Yos-1)





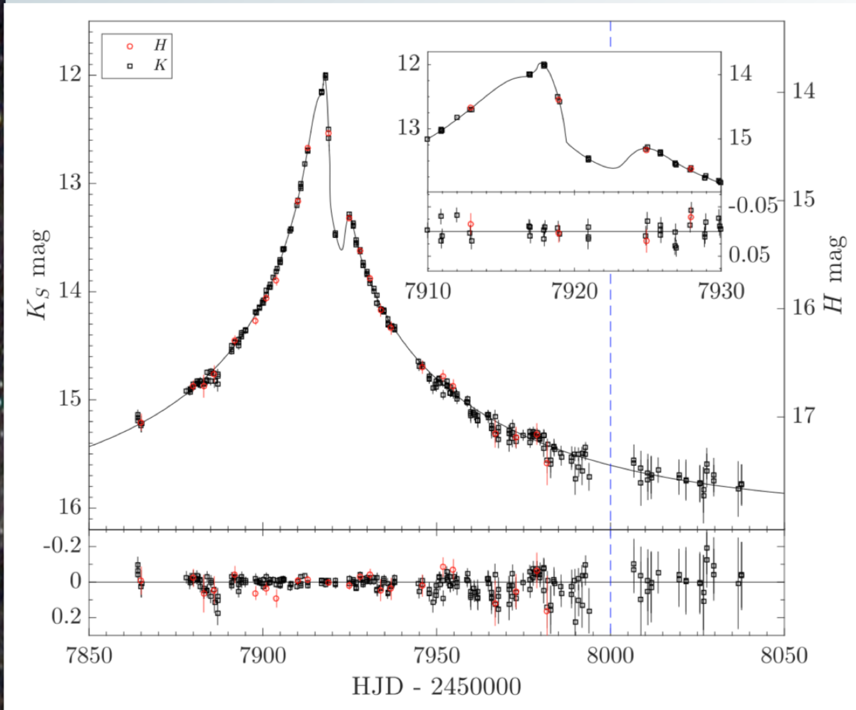
Shvartzvald et al. 2018

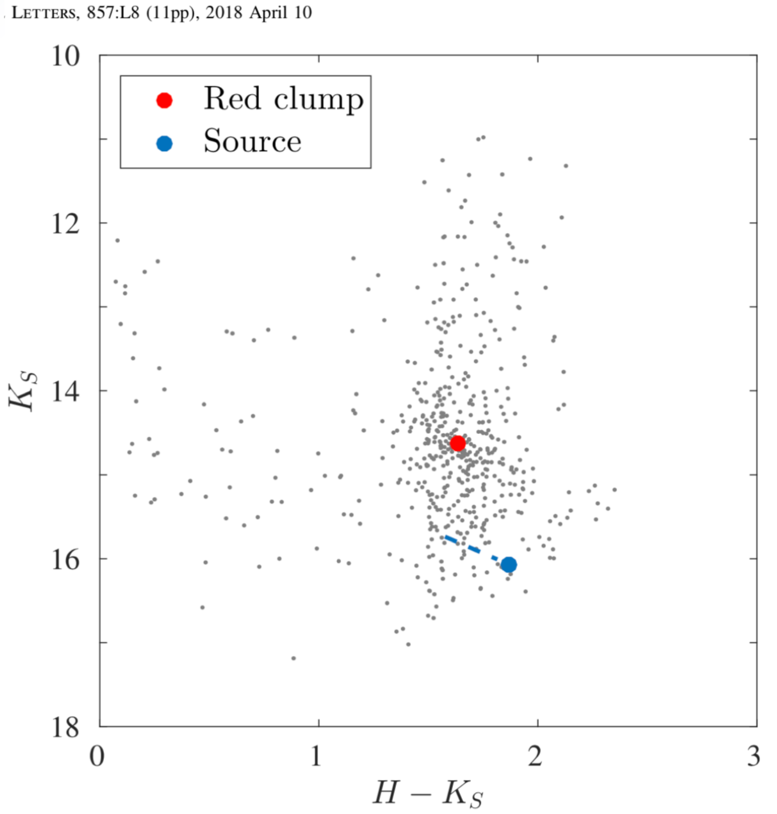
Parameter	Full dataset
t_0 [HJD']	7916.235 ± 0.018
u_0	$0.0292^{+0.0032}_{-0.0021}$
t_E [d]	$103.9^{+6.9}_{-8.8}$
$\rho[10^{-3}]$	$6.38^{+0.63}_{-0.45}$
α [rad]	$3.7004^{+0.0069}_{-0.0061}$
s	$1.0325^{+0.0029}_{-0.0032}$
q $[10^{-3}]$	$1.44^{+0.15}_{-0.10}$
K_s	16.09 ± 0.09
f_b	$0.26^{+0.21}_{-0.29}$
$(H-K)_s$	1.810 ± 0.003
$t_{\rm eff}$ [d]	3.041 ± 0.032
t_* [d]	0.6624 ± 0.0078
qt_E [d]	0.1498 ± 0.0031
f_{lim}	199.2 ± 1.2

M_1 $[M_\odot]$	$0.81^{+0.31}_{-0.34}$
M_2 $[M_J]$	$1.22^{+0.48}_{-0.51}$
$r_\perp$ [AU]	$3.24^{+0.83}_{-0.91}$
D_L [kpc]	$5.4^{+1.4}_{-1.7}$

Differential Extinction

Differential extinction reduces the accuracy of source star determination, e.g. UKIRT-2017-BLG-001Lb (Shvartzvald+ 2018),
with clump color dispersion = 0.16 mag (cf. 0.04 mag intrinsic) and clump magnitude dispersion = 0.35 mag (cf. 0.17 mag intrinsic)





Summary

The UKIRT microlensing survey is serving as a precursor for WFIRST by

- mapping the microlensing event rate
- identifying regions with high differential extinction
- developing analysis tools (e.g. machine learning)
- enabling hands-on experience for new microlensers (e.g. me)

Beyond microlensing, the survey data is available for

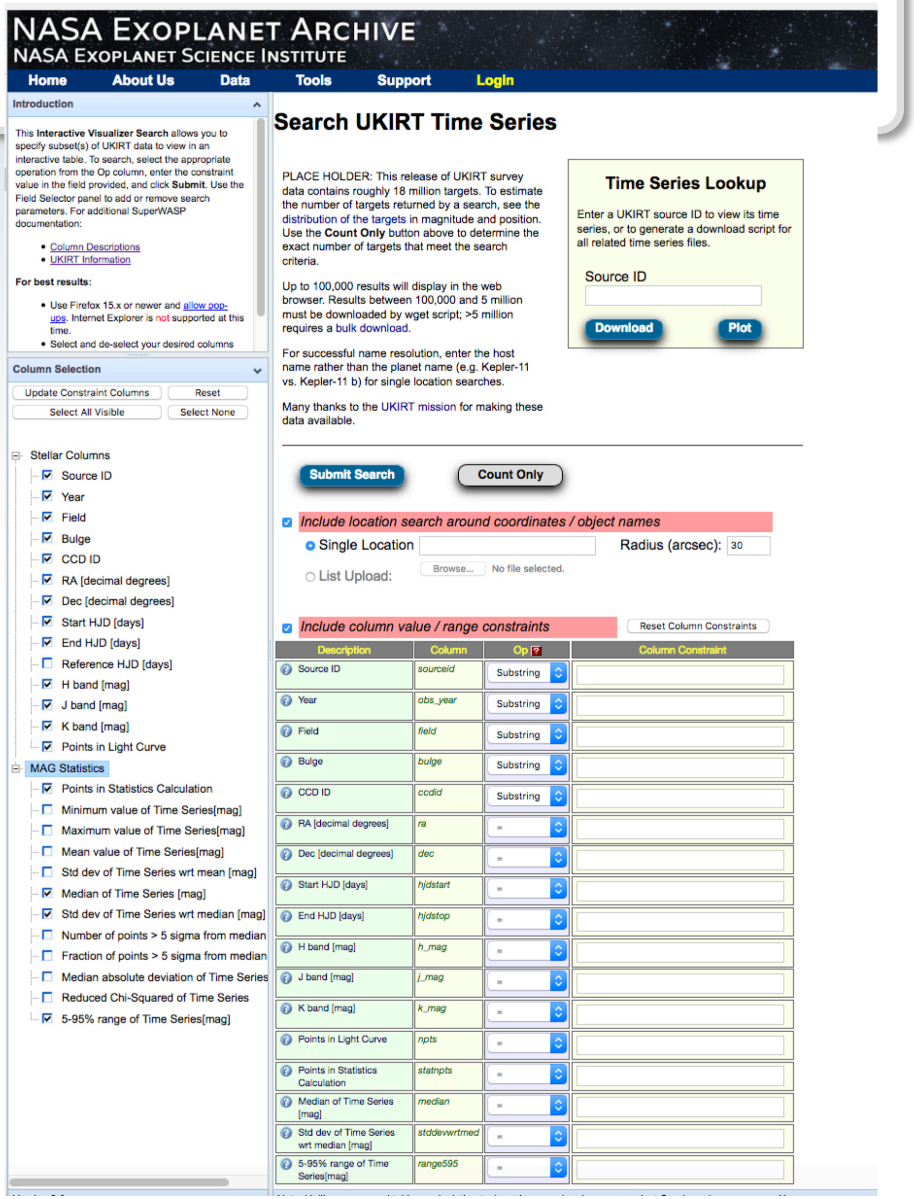
- variable stars
- outlier events
- extinction maps
- Galactic structure
- etc.

Public Lightcurves

2015-2018 lightcurves are available online
(78M independent lightcurves for ~35M sources)

Standard NExScI archive tools for selection, sorting, visualization, etc.

Flagging of known events; cross-matching with OGLE/MOA microlensing surveys.



This **Interactive Visualizer Search** allows you to specify subset(s) of UKIRT data to view in an interactive table. To search, select the appropriate operation from the Op column, enter the constraint value in the field provided, and click **Submit**. Use the Field Selector panel to add or remove search parameters. For additional SuperWASP documentation:

- Column Descriptions
- UKIRT Information

For best results:

- Use Firefox 15.x or newer and allow pop-ups. Internet Explorer is not supported at this time.
- Select and de-select your desired columns

Stellar Columns
- ☑ Source ID
- ☑ Year
- ☑ Field
- ☑ Bulge
- ☑ CCD ID
- ☑ RA [decimal degrees]
- ☑ Dec [decimal degrees]
- ☑ Start HJD [days]
- ☑ End HJD [days]
- ☐ Reference HJD [days]
- ☑ H band [mag]
- ☑ J band [mag]
- ☑ K band [mag]
- ☑ Points in Light Curve

MAG Statistics
- ☑ Points in Statistics Calculation
- ☐ Minimum value of Time Series[mag]
- ☐ Maximum value of Time Series[mag]
- ☐ Mean value of Time Series[mag]
- ☐ Std dev of Time Series wrt mean [mag]
- ☑ Median of Time Series [mag]
- ☑ Std dev of Time Series wrt median [mag]
- ☐ Number of points > 5 sigma from median
- ☐ Fraction of points > 5 sigma from median
- ☐ Median absolute deviation of Time Series
- ☐ Reduced Chi-Squared of Time Series
- ☑ 5-95% range of Time Series[mag]

Search UKIRT Time Series

PLACE HOLDER: This release of UKIRT survey data contains roughly 18 million targets. To estimate the number of targets returned by a search, see the distribution of the targets in magnitude and position. Use the **Count Only** button above to determine the exact number of targets that meet the search criteria.

Up to 100,000 results will display in the web browser. Results between 100,000 and 5 million must be downloaded by wget script; >5 million requires a bulk download.

For successful name resolution, enter the host name rather than the planet name (e.g. Kepler-11 vs. Kepler-11 b) for single location searches.

Many thanks to the UKIRT mission for making these data available.

Time Series Lookup

Enter a UKIRT source ID to view its time series, or to generate a download script for all related time series files.

Source ID

☑ *Include location search around coordinates / object names*

● Single Location   Radius (arcsec): 30

○ List Upload: No file selected.

☑ *Include column value / range constraints*

Description	Column	Op	Column Constraint
Source ID	sourceid	Substring	
Year	obs_year	Substring	
Field	field	Substring	
Bulge	bulge	Substring	
CCD ID	ccdid	Substring	
RA [decimal degrees]	ra	=	
Dec [decimal degrees]	dec	=	
Start HJD [days]	hjdstart	=	
End HJD [days]	hjdstop	=	
H band [mag]	h_mag	=	
J band [mag]	j_mag	=	
K band [mag]	k_mag	=	
Points in Light Curve	npts	=	
Points in Statistics Calculation	statnpts	=	
Median of Time Series [mag]	median	=	
Std dev of Time Series wrt median [mag]	stddevwrtmed	=	
5-95% range of Time Series[mag]	range595	=	

https://exoplanetarchive.ipac.caltech.edu/docs/UKIRTMission.html
https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblSearch/nph-tblSearchInit?app=ExoTbls&config=ukirttimeseries

backup…

Scientific goals of UKIRT survey

NIR event rate as a function of (*l*,*b*):

- Crucial for *WFIRST* field optimization
- Combined with dust models → Galactic structure

Event timescale as a function of (*l*,*b*):

- Bulge-bulge events are expected to be shorter (Gould 1995)

NIR coverage of events:

- Source color - for Einstein radius (with finite source effects)
- NIR source flux - for future AO lens flux measurements

New science:

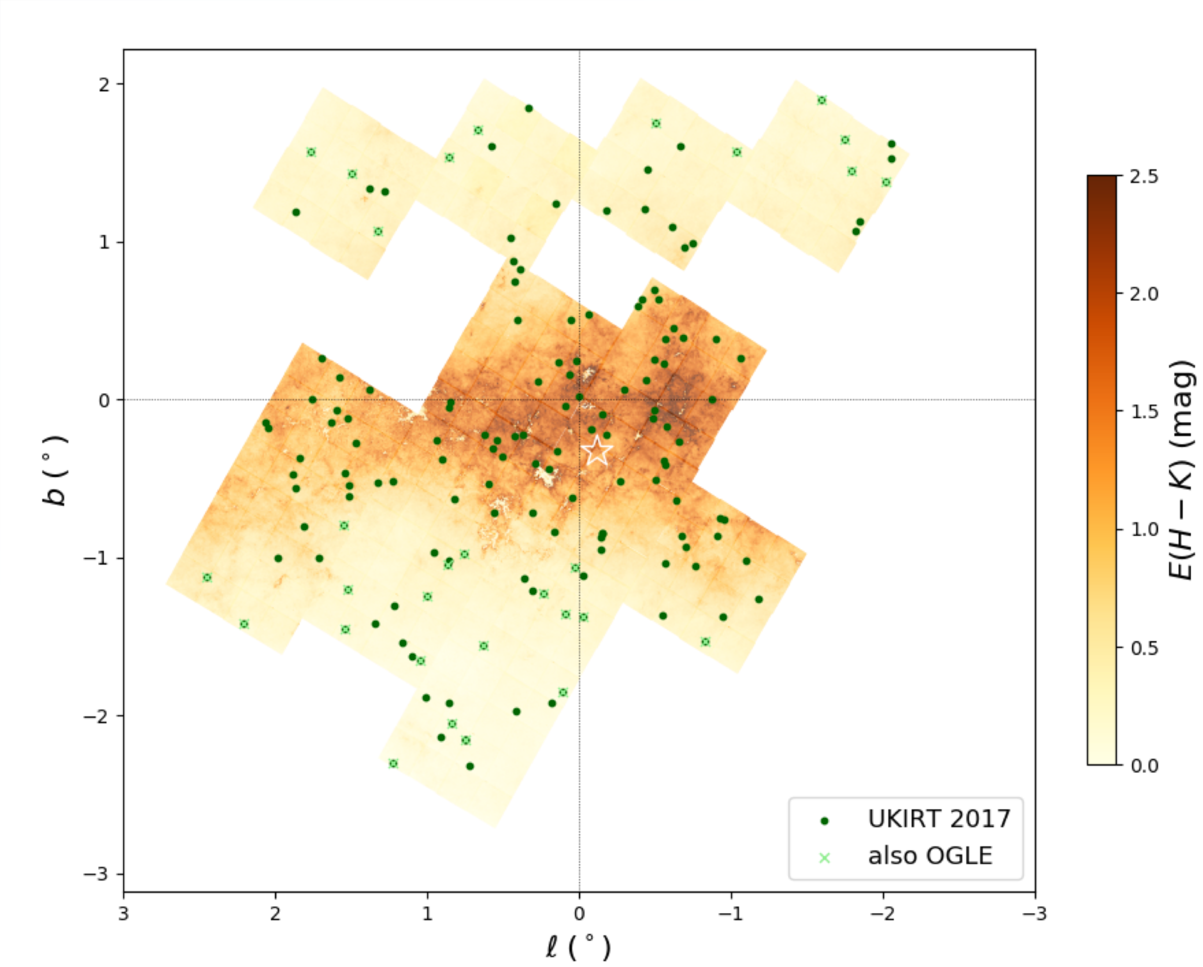
- High cadence (daily) observations of unexplored regions (Galactic center).

Target Field Selection?

1. data reduction
2. microlensing event identification
3. completeness correction
4. updated galactic model / incorporation into mission yield simulations
5. WFIRST target field selection
    - **event rate map**
    - **differential extinction**
    - overlap with ground-based surveys / potential ground-based follow-up

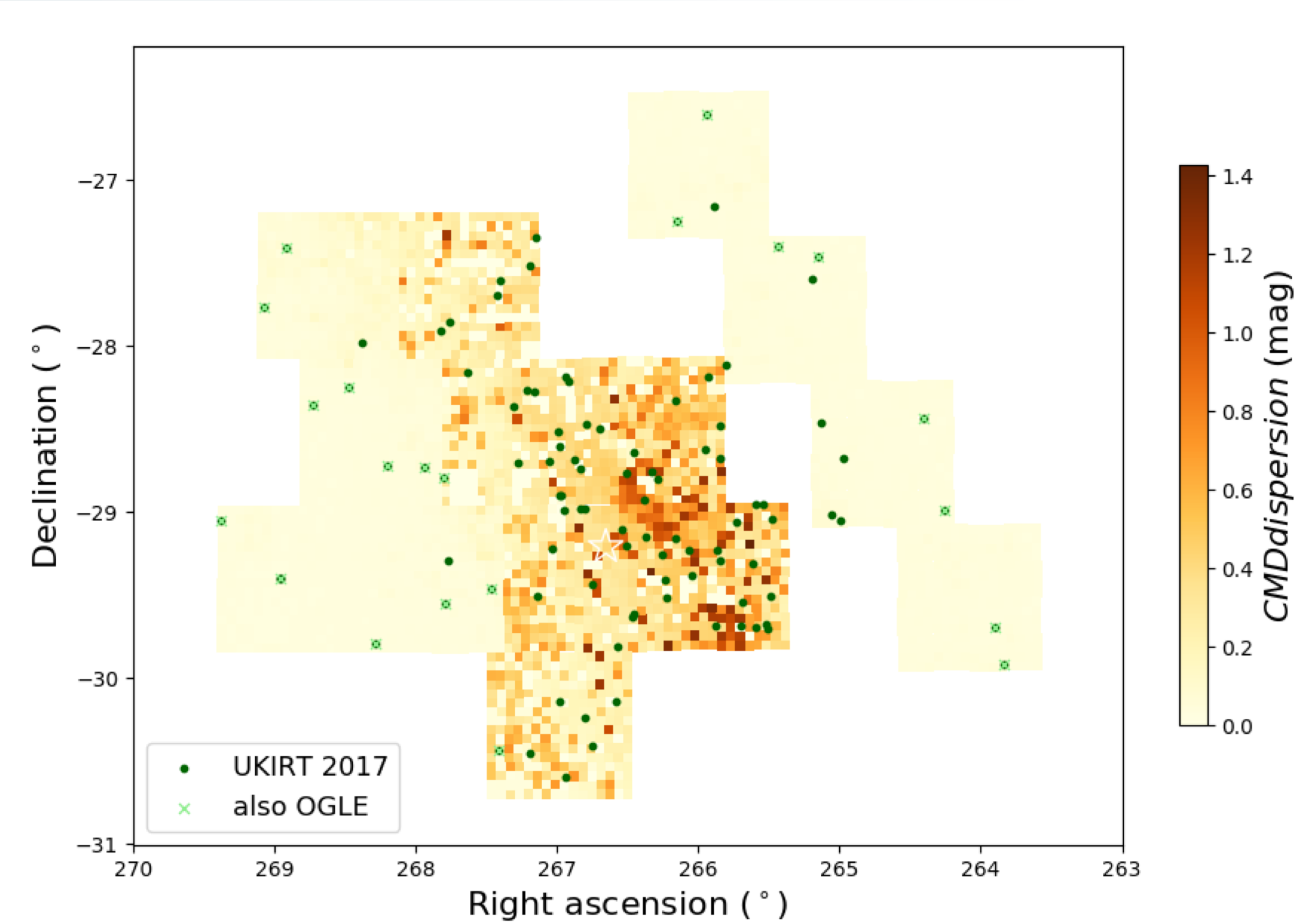
Extinction

- Median H-K reddening within 20-arcsec pixels.
- High extinction is a problem for visible-light observations.



Differential Extinction

- Based on CMD within 3-arcmin pixels,
  using David Nataf's code to find spread of the red clump.

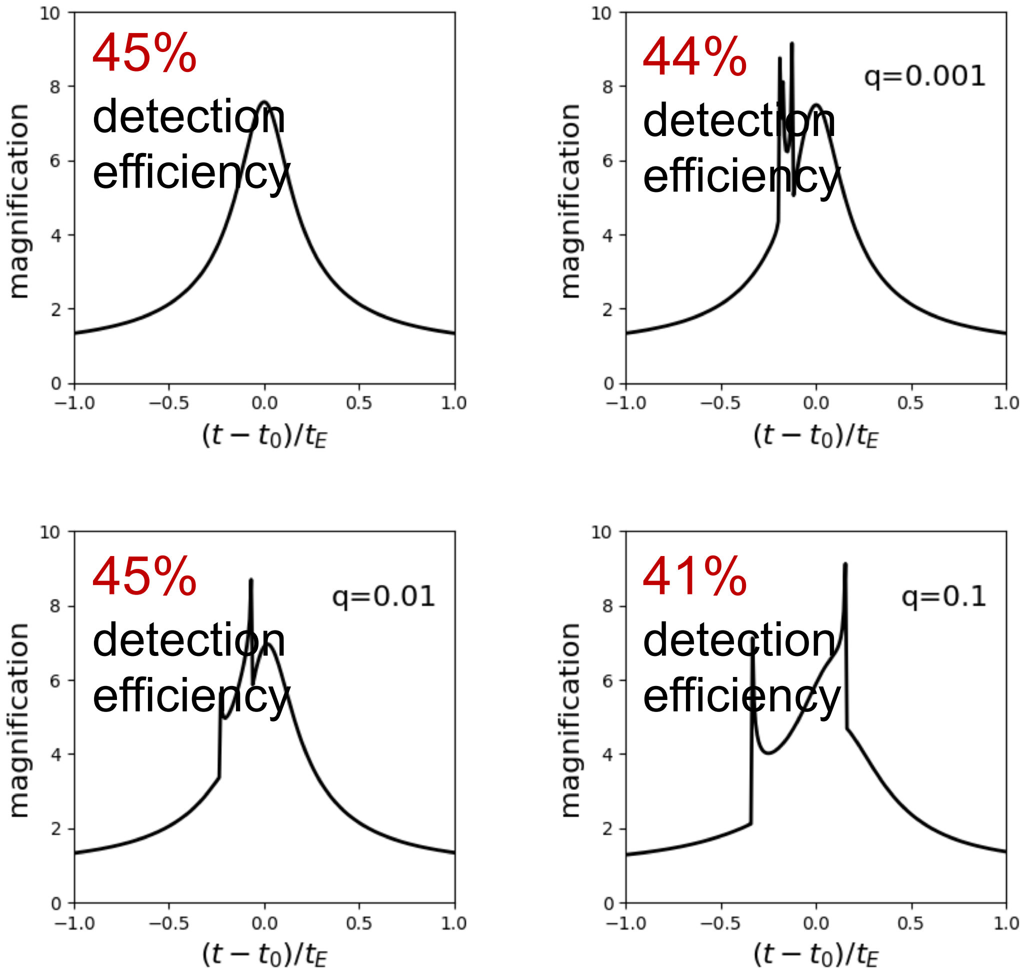


Application to Binaries

The machine learning classifier has been optimized to identify single-lens-single-source events. *What about binaries?*

Detection efficiency decreases for highly anomalous light curves, but not by much - **~10% reduction for q=0.1**

Most importantly, this selection effect can be quantified and corrected.



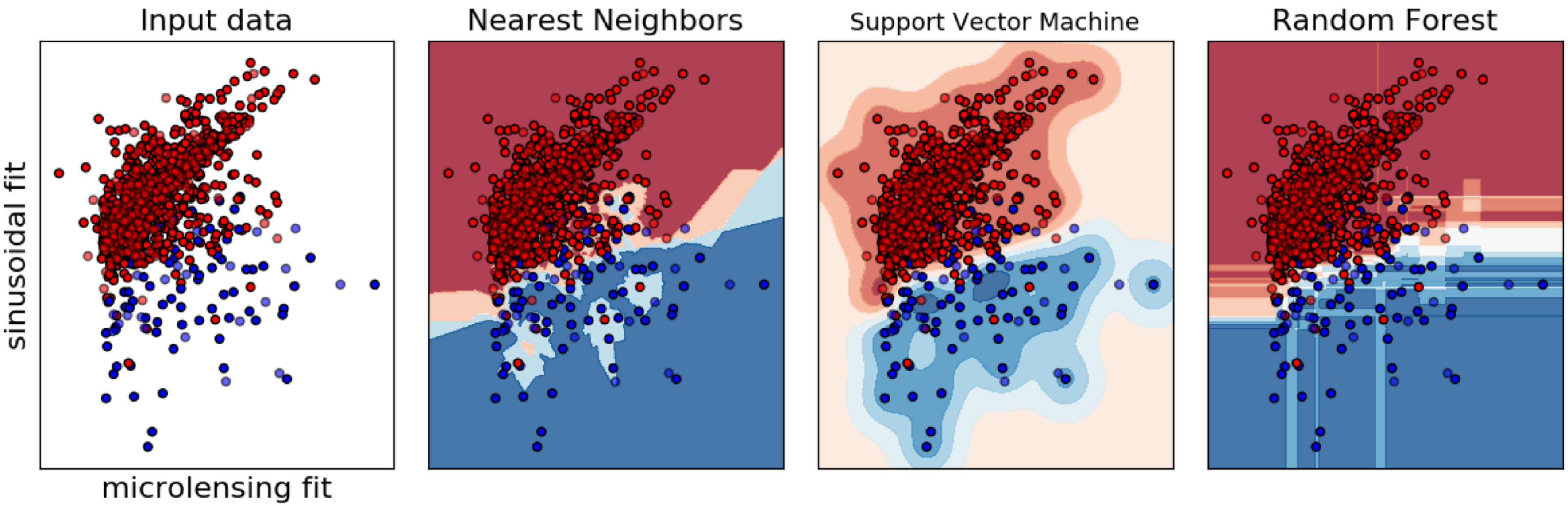
Step 3: Pick a Classifier

A complicated N-dimensional space has to somehow be divided into different categories (microlensing, variable, artifact, etc).

There's lots of options here:

nearest neighbor, support vector machine, gaussian process, decision tree, random forest, neural net, AdaBoost, naive Bayes,...

We find that **random forest** is fast and accurate within our training set.



blue=microlensing events; red=other variables

Step 2: Feature Selection

Class: Movies you like

Movie features:

- length of movie name
- director
- genre

Class: Microlensing event

Light curve features:

- flux, dispersion
- # of obs., # of successful obs.
- χ^2 for microlens curve, flat line, sloped line, sinusoid, drop 1 or 2 pts.
- $\Delta\chi^2$ between those fits
- fit parameters and error bars for each fit
- time of event relative to observing window
- random numbers (for testing purposes)

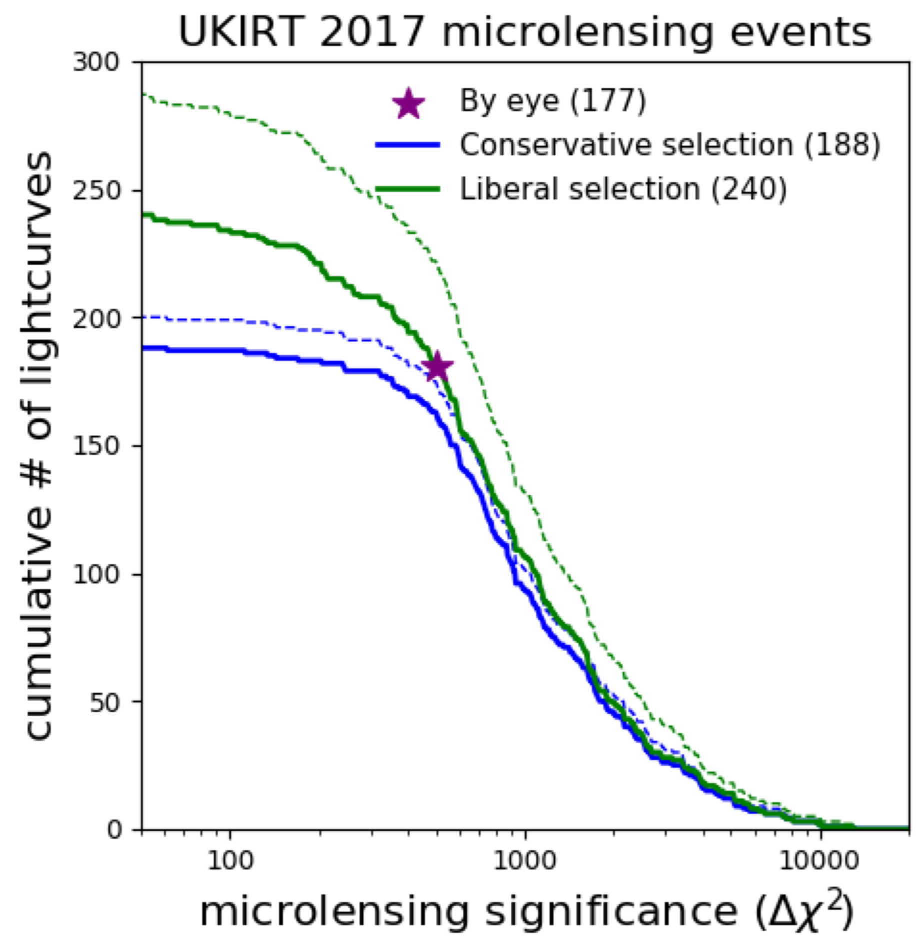
→ The classifier will tell you the ranked importance of each.

Goal 1: Efficiently Detect Events

How well do we do, compared to by-eye selection?

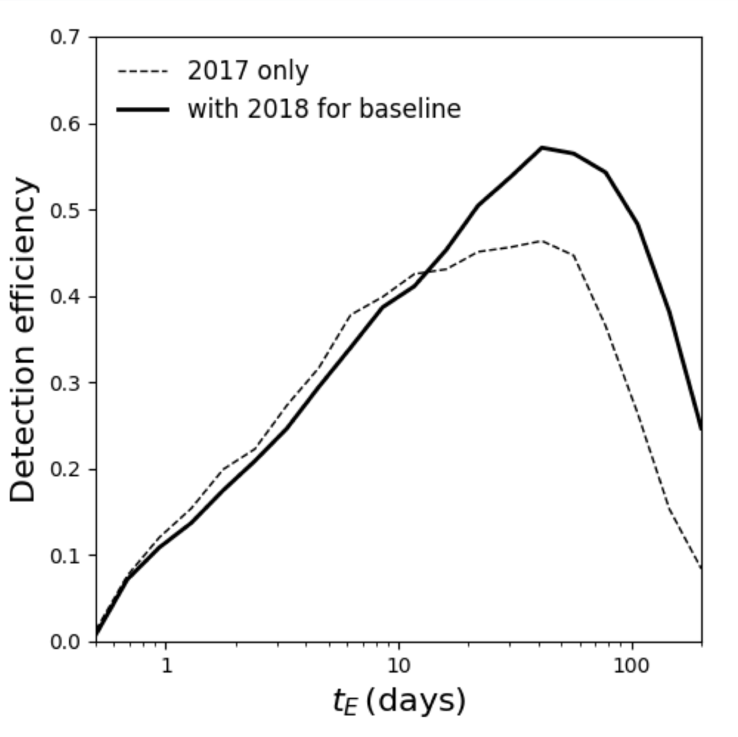
The random-forest classifier provides a microlensing probability for each lightcurve. These probabilities are not well calibrated, but they are meaningful relative to each other.

Normally we count an event as microlensing if the probability is >50%.
By varying this threshold, we can make a more conservative or more liberal selection.

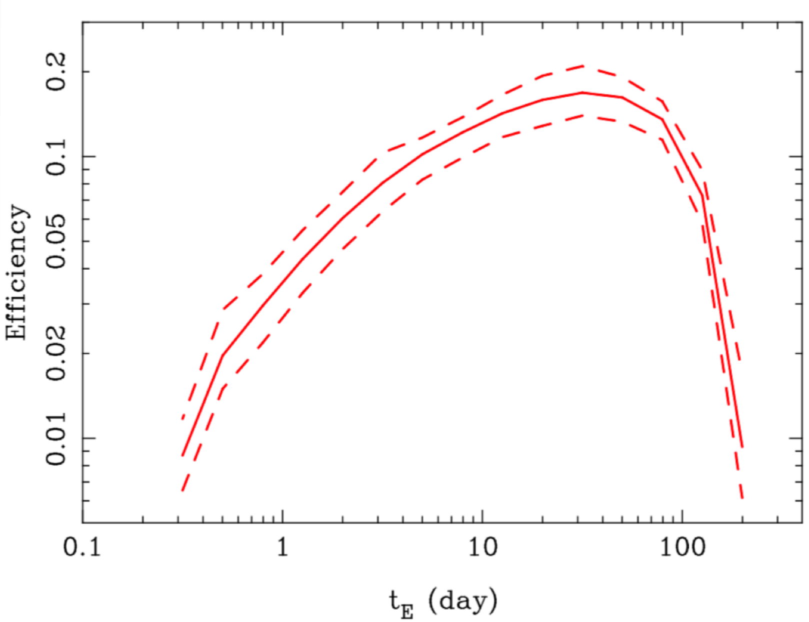


Machine learning vs traditional selection

machine learning can give
40-60% detection efficiency



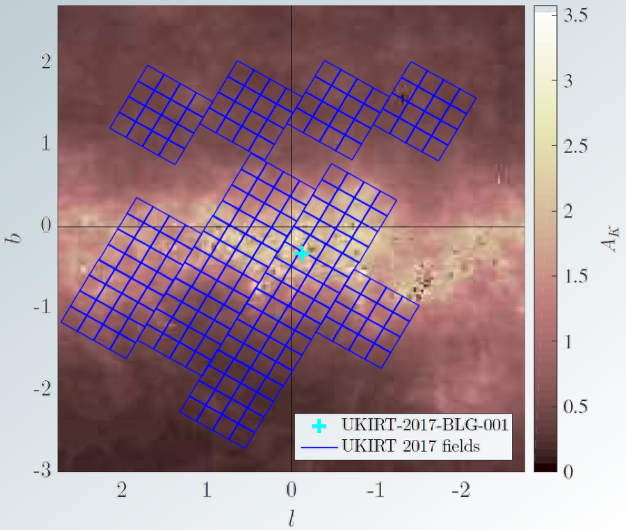
an example of deterministic cuts
with 10-20% detection efficiency

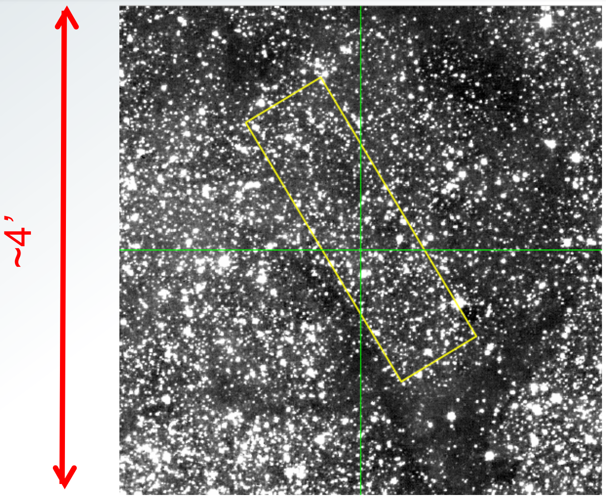


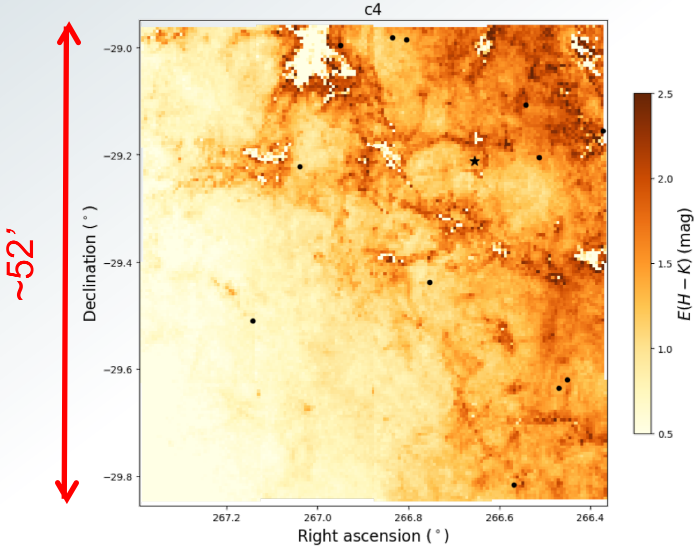
MOA survey; Sumi et al. 2011

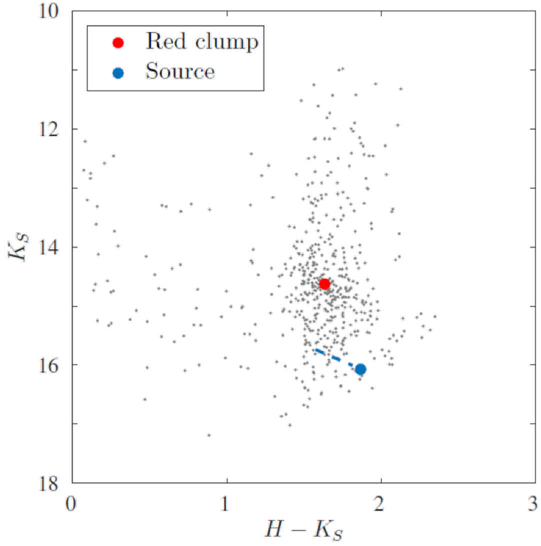
It is impossible to do a self-consistent comparison with OGLE/MOA. Instead, we are working on a UKIRT vs UKIRT comparison, machine learning vs deterministic cuts. (w/ advice from Przemek).

UKIRT-2017-BLG-001: Extinction









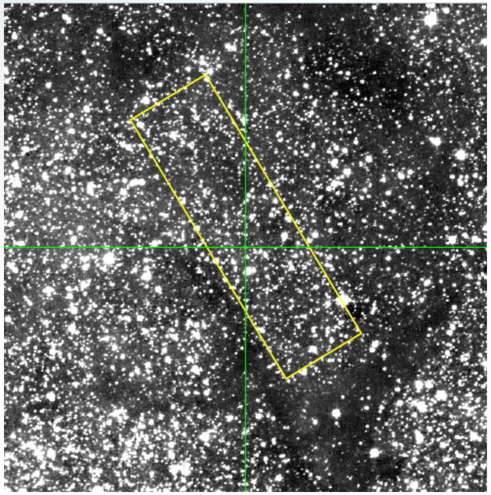
UKIRT-2017-BLG-001: Extinction

Lessons for *WFIRST*

- Red clump color dispersion:

$$\sigma_{(H-K_S)}=0.16$$

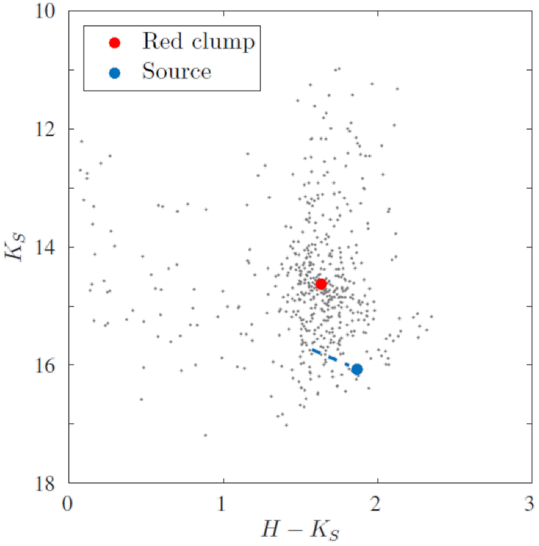
  - Intrinsic = 0.04
  - Reddening = 0.15
- Dust scale height = 120pc (0.86°@8kpc)



- Red clump magnitude dispersion:

$$\sigma_{K_S}=0.36$$

  - Intrinsic = 0.17
  - Extinction = 0.17
  - DM = 0.28!
- Thin disk scale height = 300pc (2.1°@8kpc)



UKIRT-2017-BLG-001: Extinction

Lessons for *WFIRST*

- Limitations of one season
- Estimation of θ_*

  …and thus physical properties
- Far disk population

- Possible solutions:
  - Multi-band information
  - Avoid high spatial differential extinction fields
  - >1° off the plane?