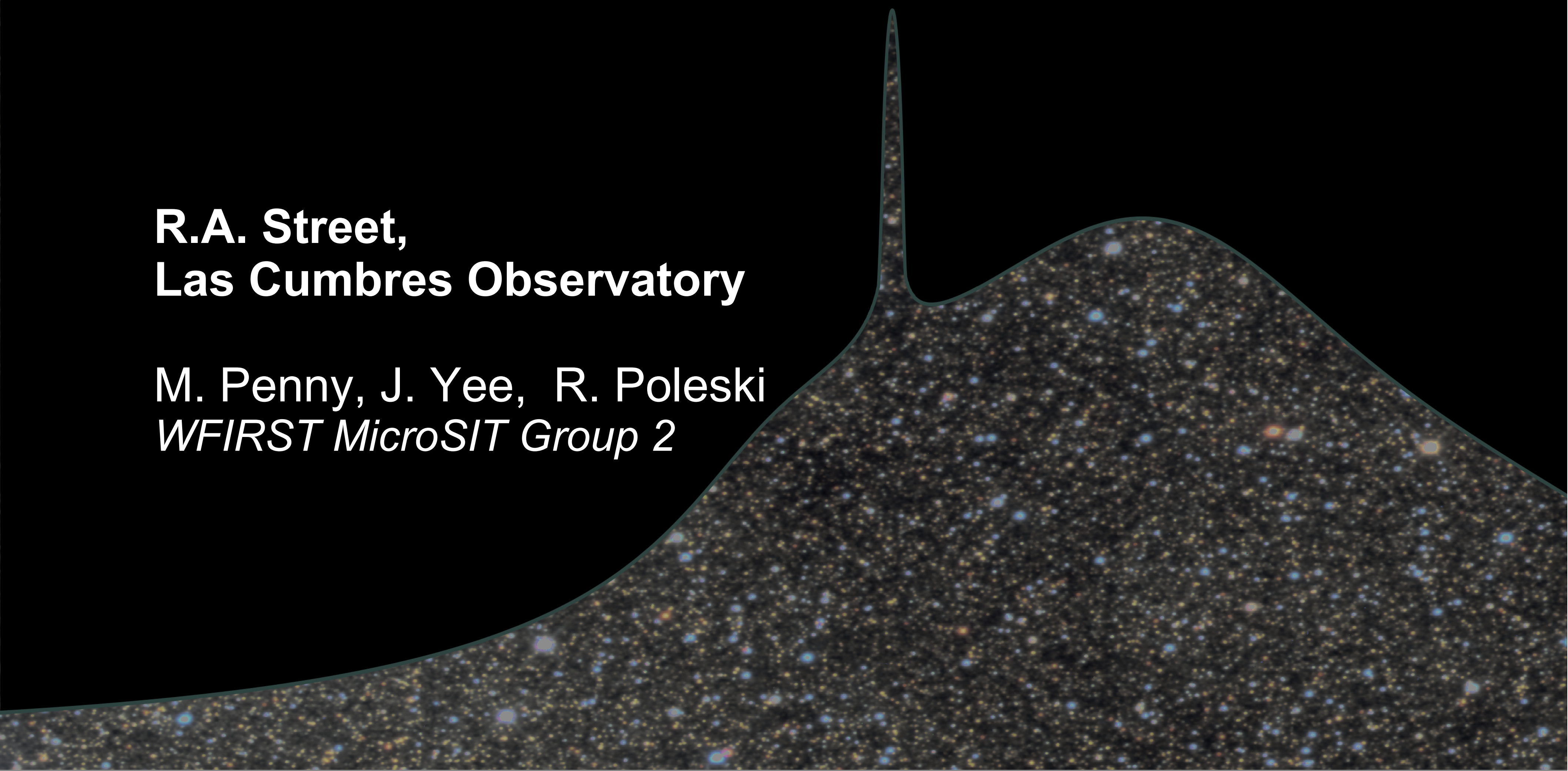
Results of Microlensing Data Challenge 1

R.A. Street,
Las Cumbres Observatory

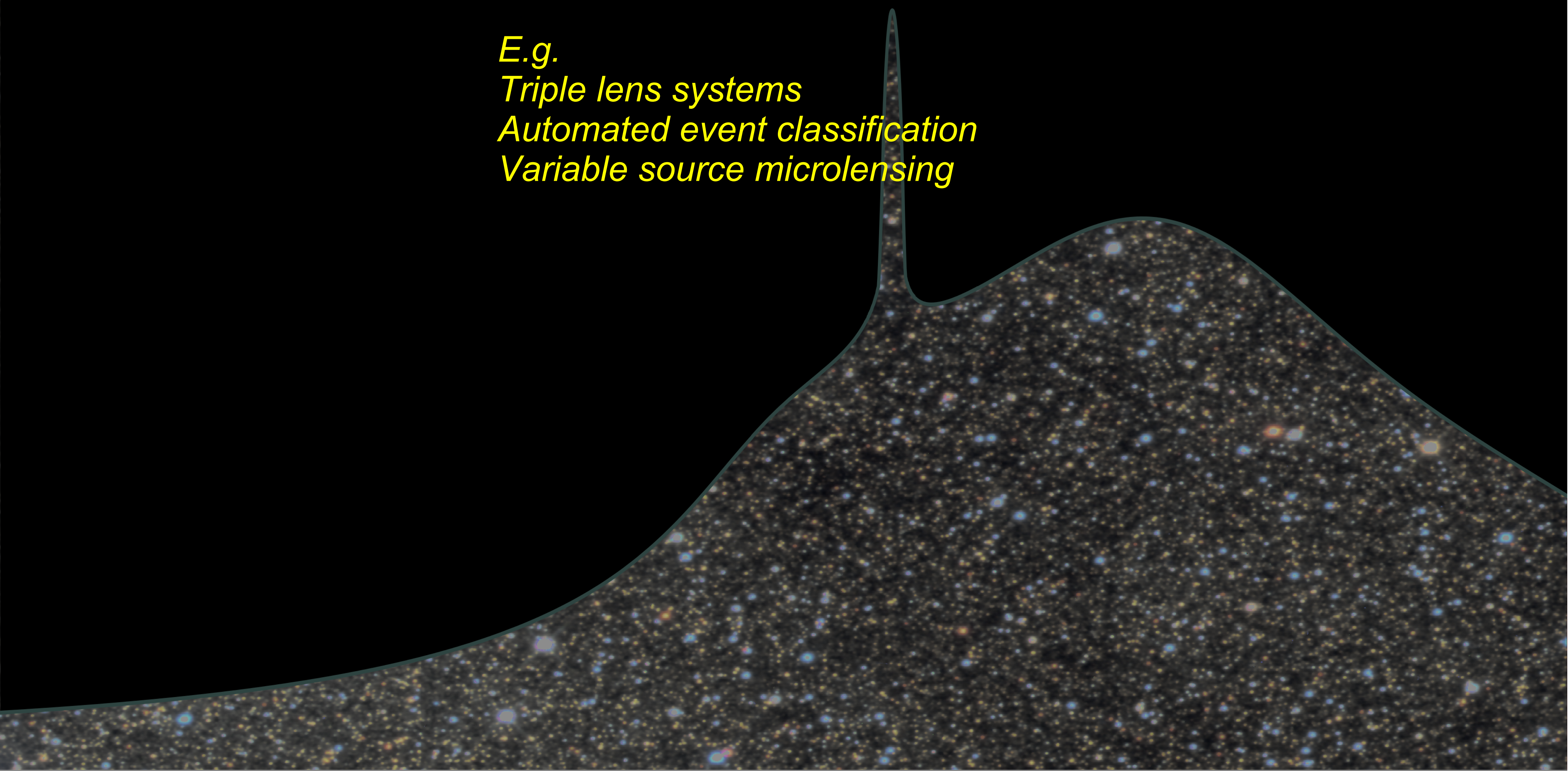
M. Penny, J. Yee, R. Poleski
WFIRST MicroSIT Group 2



Motivation

Though microlensing has been observed for a while, there are still a number of unsolved challenges with microlensing analysis

E.g.
Triple lens systems
Automated event classification
Variable source microlensing

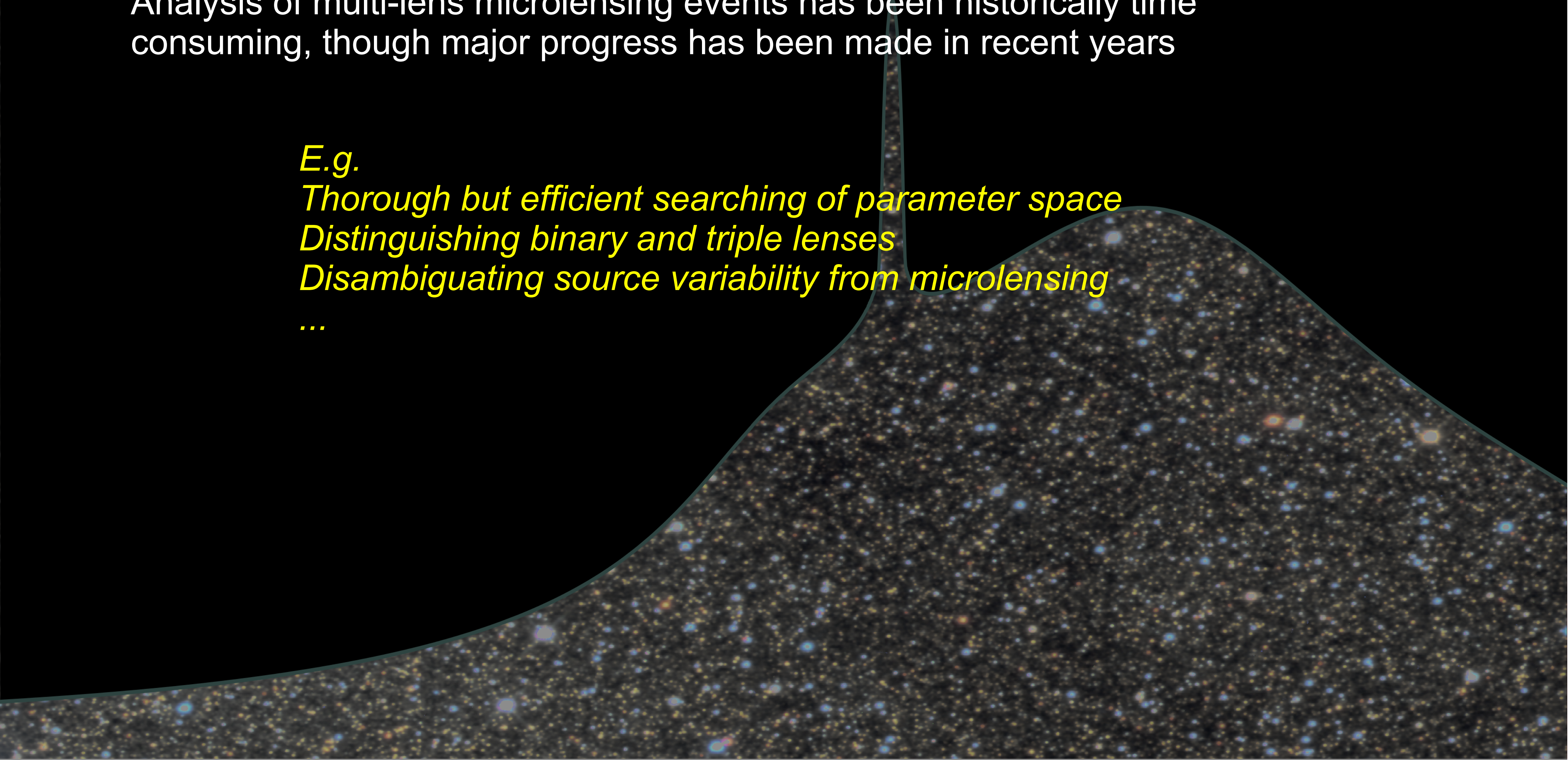


Motivation

Though microlensing has been observed for a while, there are still a number of unsolved challenges with microlensing analysis

Analysis of multi-lens microlensing events has been historically time consuming, though major progress has been made in recent years

E.g.
Thorough but efficient searching of parameter space
Distinguishing binary and triple lenses
Disambiguating source variability from microlensing
...



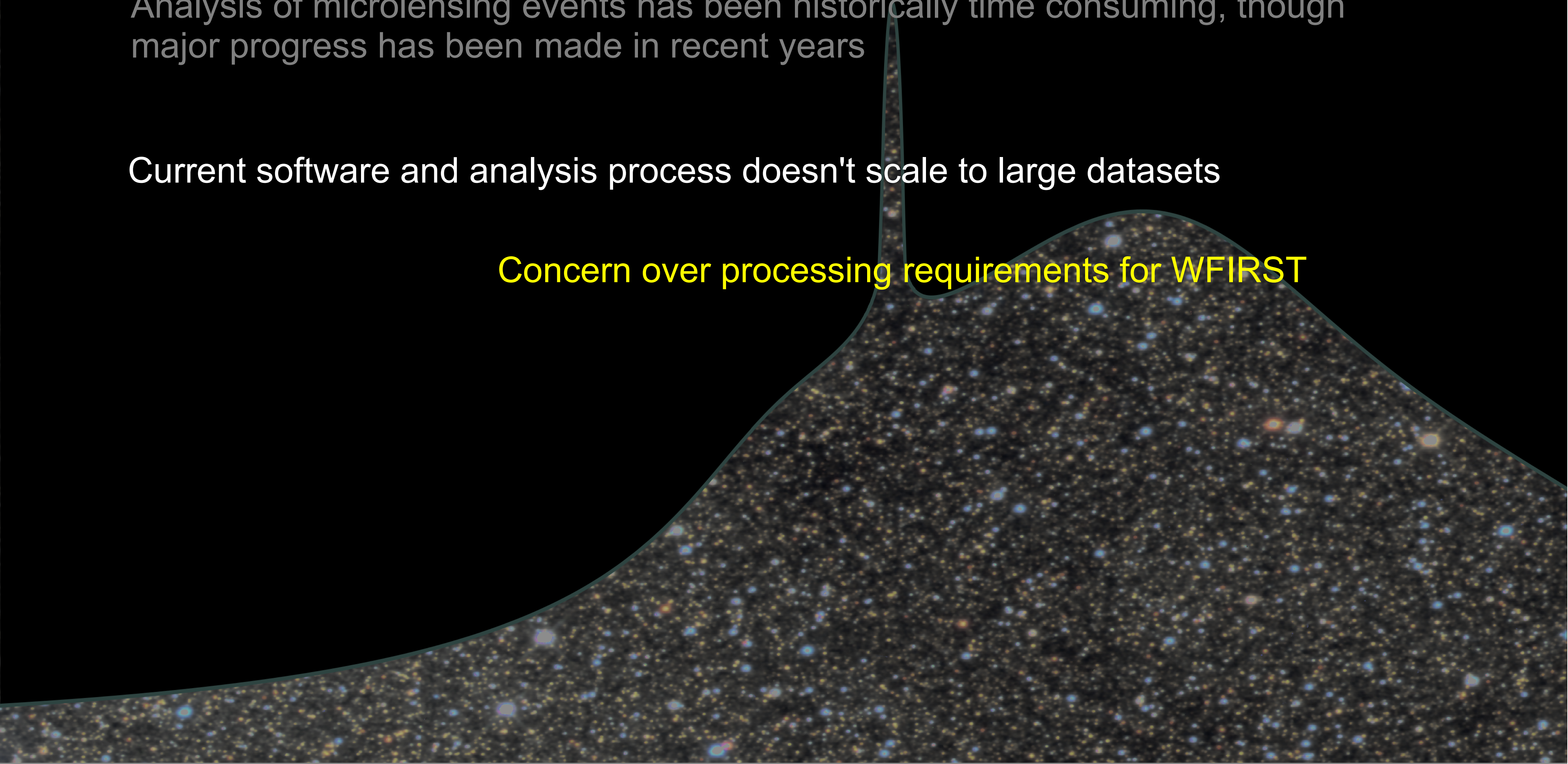
Motivation

Though microlensing has been observed for a while, there are still a number of unsolved challenges with microlensing analysis

Analysis of microlensing events has been historically time consuming, though major progress has been made in recent years

Current software and analysis process doesn't scale to large datasets

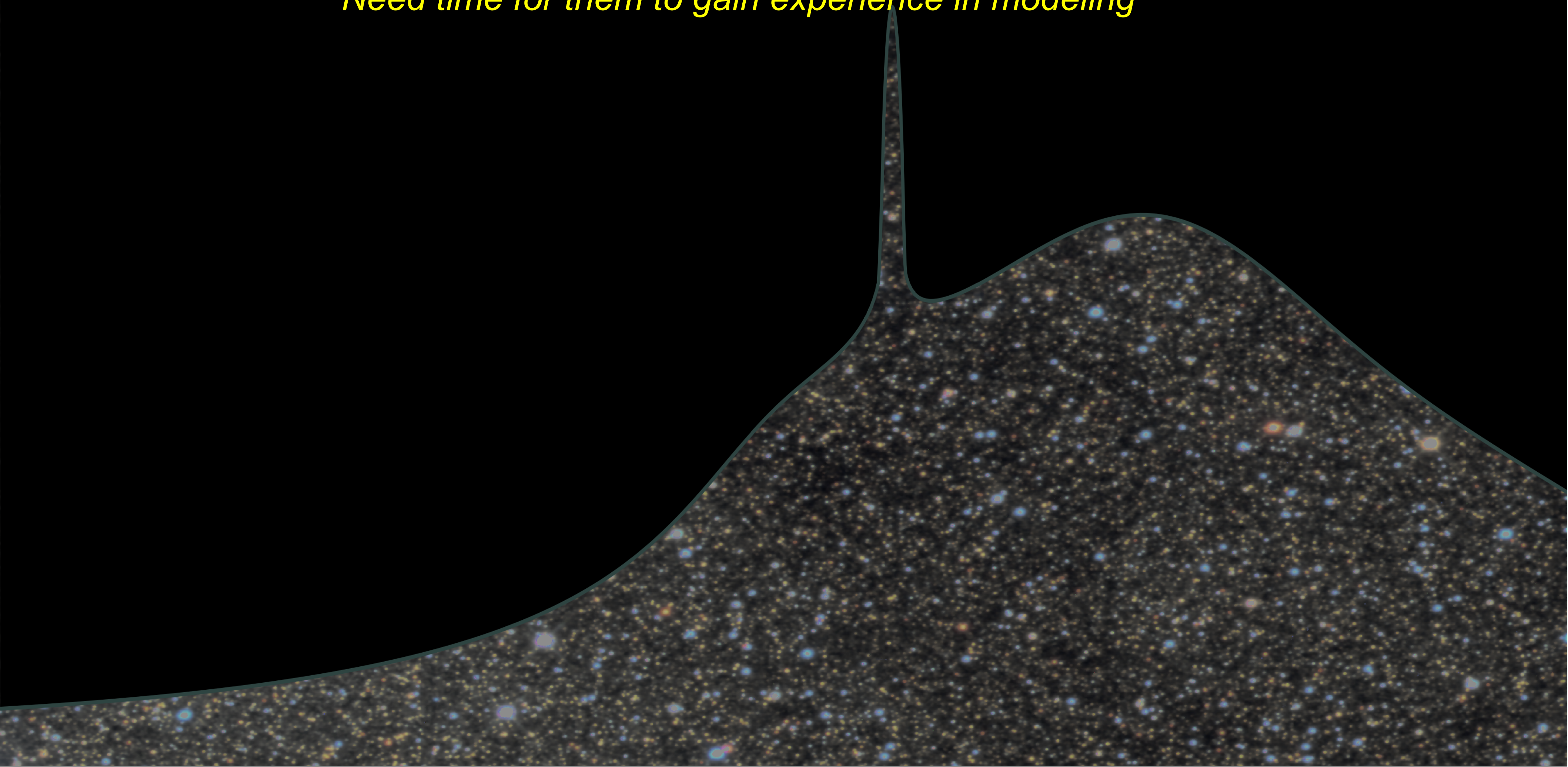
Concern over processing requirements for WFIRST



Motivation

Microlensing analysis is personnel-limited

Need to attract new people, students to the field
Need time for them to gain experience in modeling

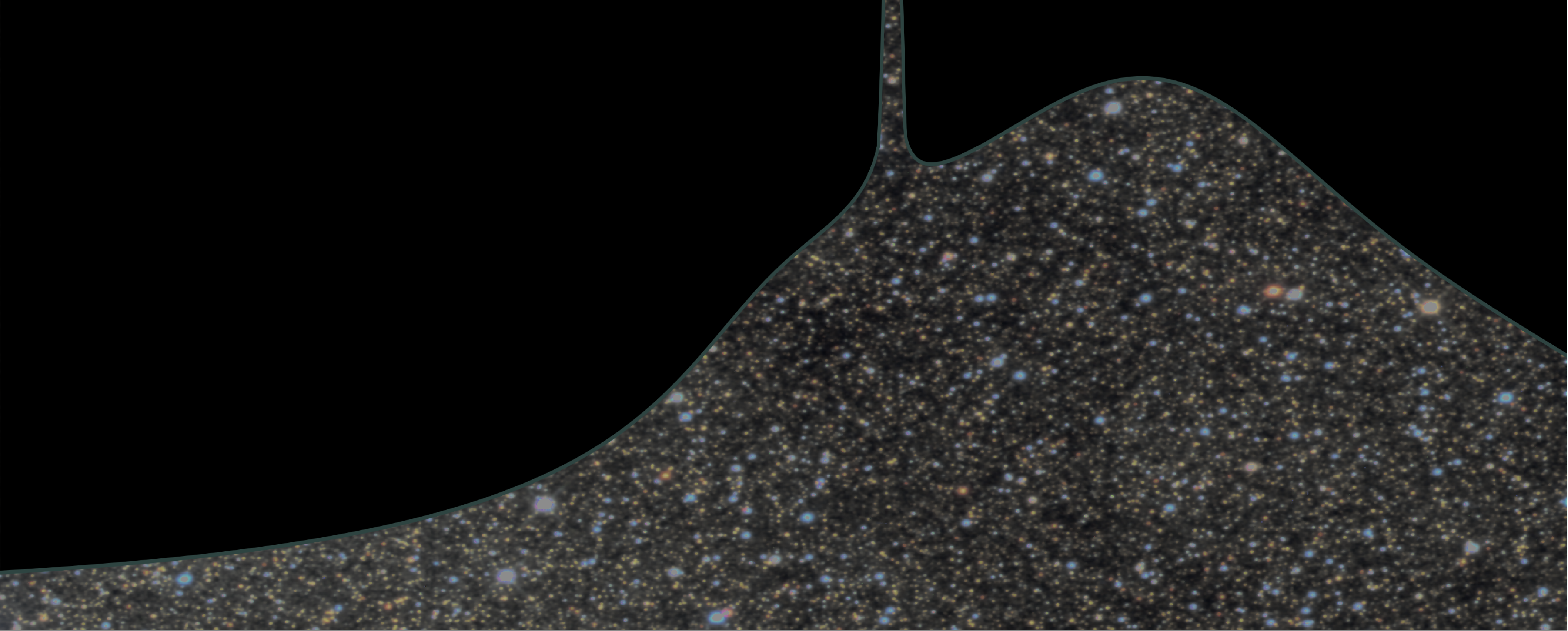


Motivation

Microlensing is personnel-limited

Would benefit from involvement of experts in mathematics, statistics, informatics

E.g.
Other fields also search for best model in large parameter space
→ Bring in new techniques



Motivation

Microlensing is personnel-limited

Would benefit from involvement of experts in mathematics, statistics, informatics

Data Challenges have been successful in stimulating engagement and innovation in other fields including exoplanets

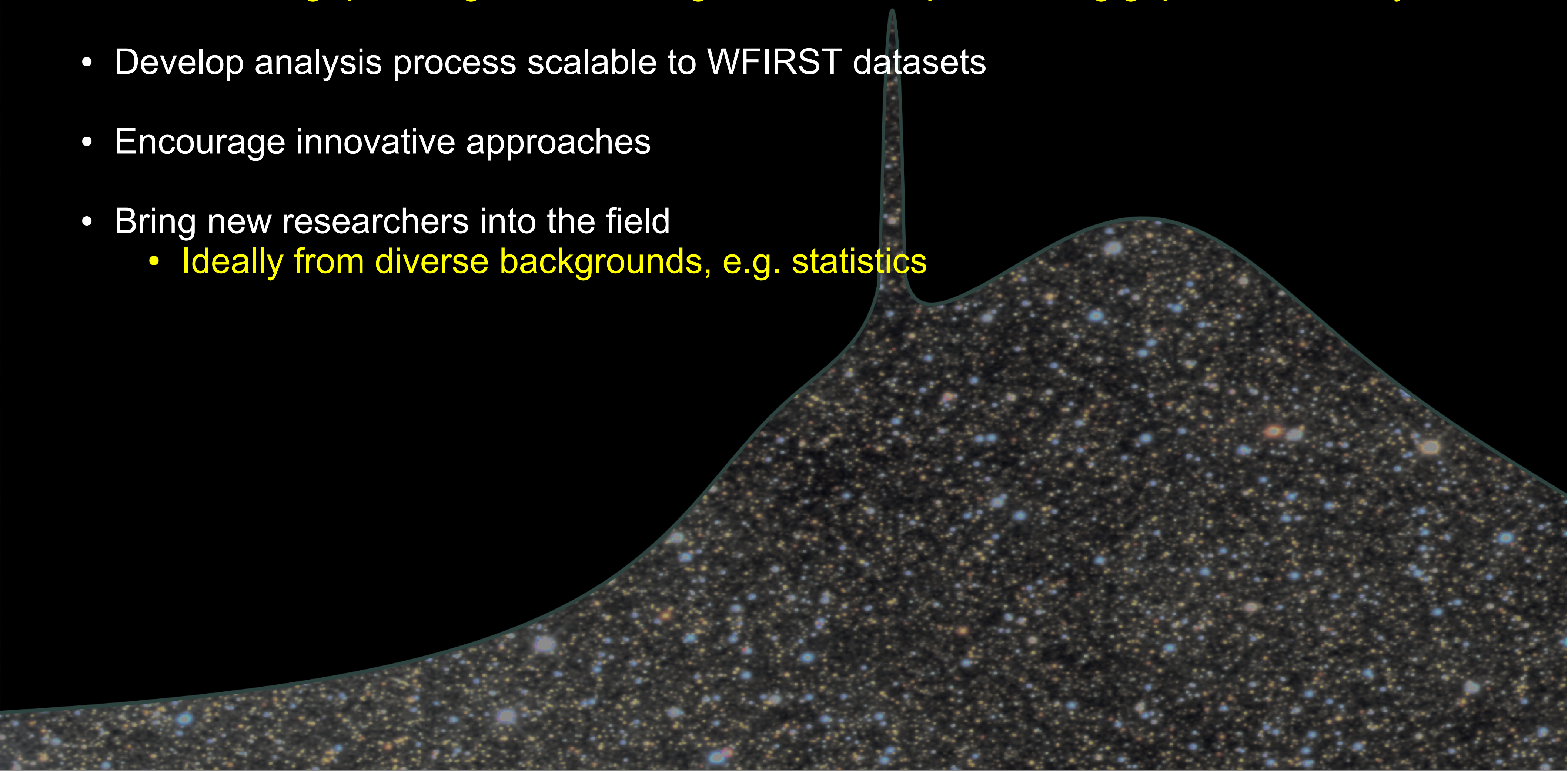
E.g.
Radial velocity fitting challenge
https://github.com/EPRV3EvidenceChallenge/Inputs
Dumusque, X. et al. (2016), A&A, 593, 5
Dumusque, X. et al. (2017), A&A, 598, 133

Transit detection
CoRoT analyses challenge

Exoplanet atmosphere spectral analysis
Hildebrant, S et al.
http://adsabs.harvard.edu/abs/2018AAS...23115803H

Goals of Data Challenge 1

- Accurately model events
- Distinguish variable stars
  - Including “pathological” cases, eg. those which peak during gaps in the survey etc.
- Develop analysis process scalable to WFIRST datasets
- Encourage innovative approaches
- Bring new researchers into the field
  - Ideally from diverse backgrounds, e.g. statistics



Simulated Data

- Matthew Penny simulated the first dataset

  Only he had access to the table of input parameters until after the submission deadline

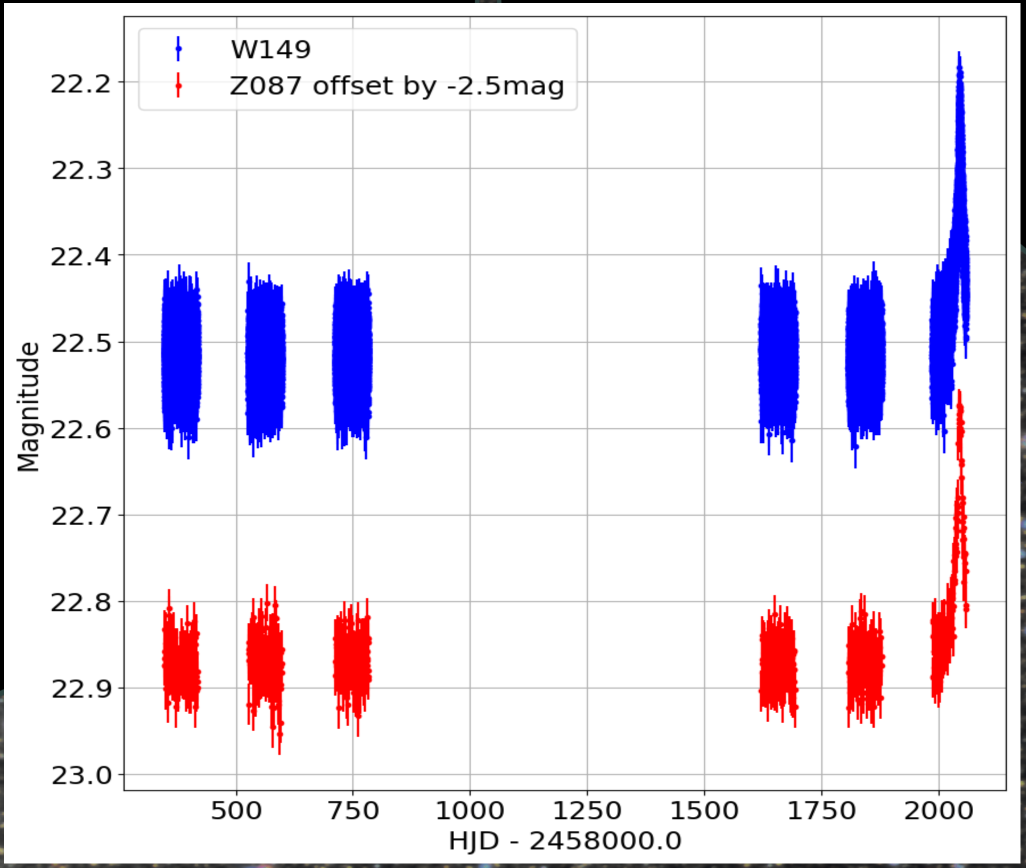
293 WFIRST lightcurves in two filters (Z087 and W149)

74 Single lenses (including FFP candidates)
83 Binary star lenses
43 Planetary binary lenses
93 Cataclysmic variables
0 Non-variables

Simulated Data

- Set of 293 lightcurves

- WFIRST lightcurves = Cadence, length and noise mimicking the nominal multi-year mission length and cadence of the Bulge survey, two filters



Entering the Challenge

http://microlensing-source.org/data-challenge

Mailing list: microlensing-data-challenge@lco.global

Github organization: https://github.com/microlensing-data-challenge

Microlensing Source

Overview | Learning | Glossary | Resources | Interactive | Opportunities | Software | Data Challenge

Microlensing Data Challenge

The analysis and modeling of microlensing events has always been a computationally-intensive and time-consuming task, requiring a powerful computer cluster as well as well sampled lightcurves. While the number of interesting events with adequate data remained fairly low, it has been practical to perform a careful interactive analysis of each one, often with the aid of a powerful computer cluster. Even so, a number of challenges remain, particularly concerning the analysis of triple lenses.

This is expected to change with next-generation surveys, especially with the launch of WFIRST. This mission is expected to detect thousands of microlensing events, including hundreds of planetary events. Clearly, our analysis techniques need an upgrade to fully exploit this dataset, and we encourage people from outside the current microlensing community to bring in diverse expertize.

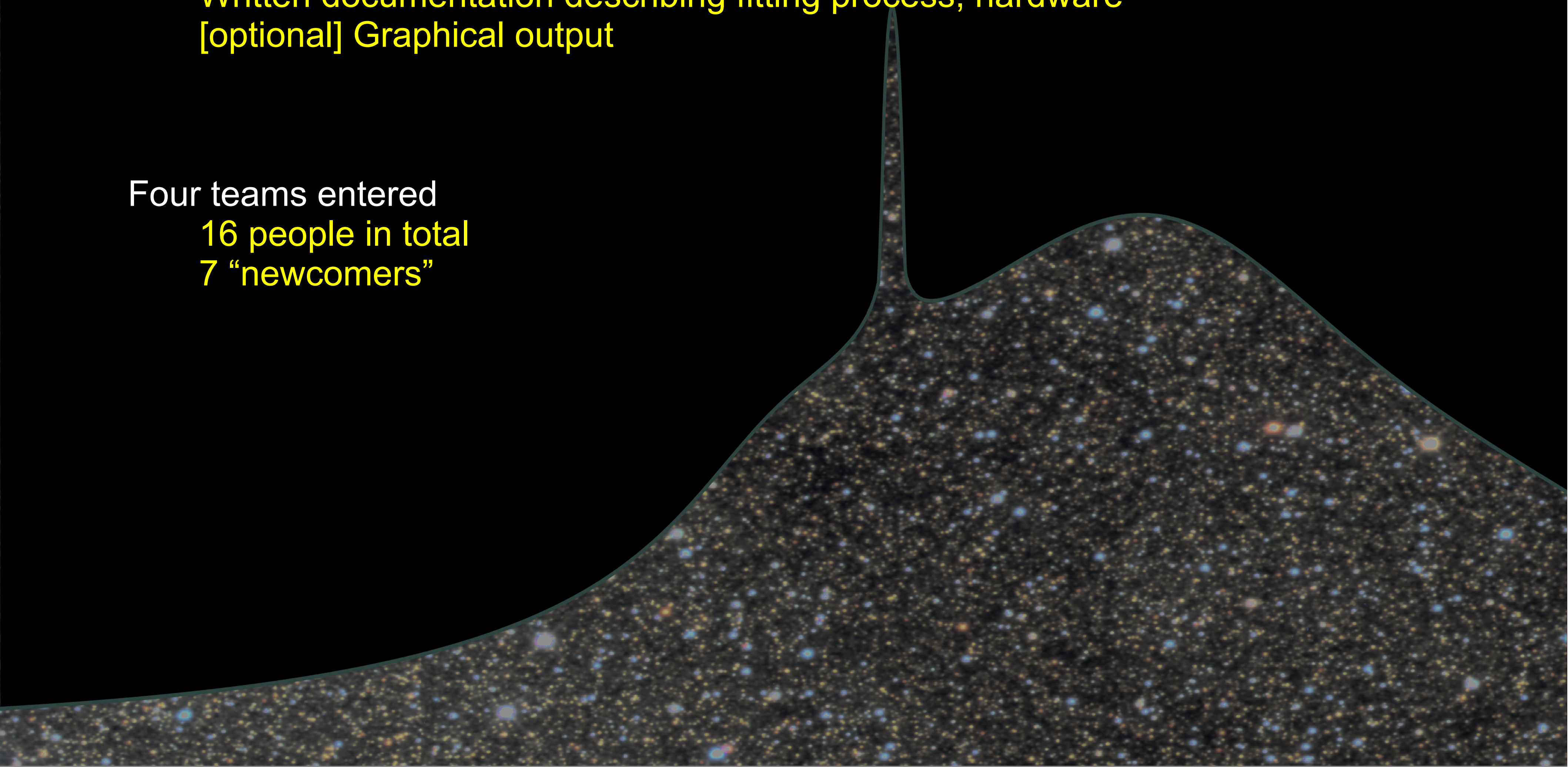
Evaluating Challenge Entries

Entry contents:

- Team contact details
- Table of fitted model parameters for each star
- Written documentation describing fitting process, hardware
- [optional] Graphical output

Four teams entered

- 16 people in total
- 7 “newcomers”



Evaluating Challenge Entries

All entries were anonymized

4-person evaluation panel + non-voting chair (RAS):

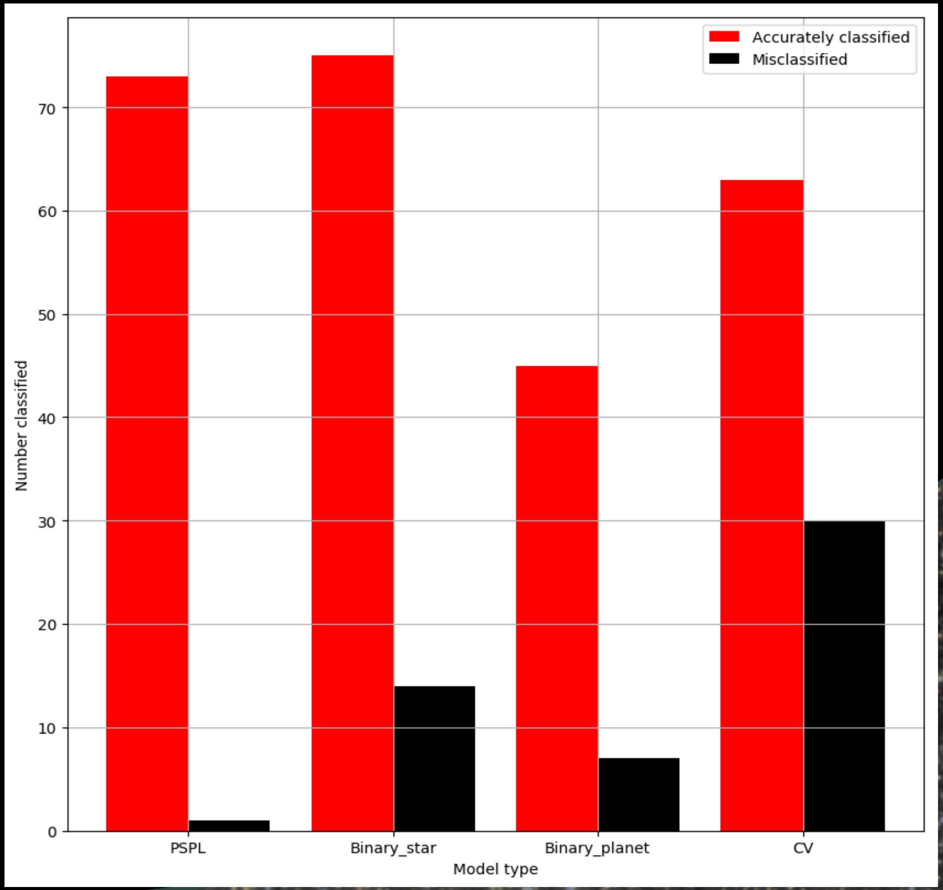
- Rachel Akeson, IPAC
- Scott Gaudi, Ohio State
- Hyungsuk Tak, Harvard
- Eamonn Kerins, Manchester

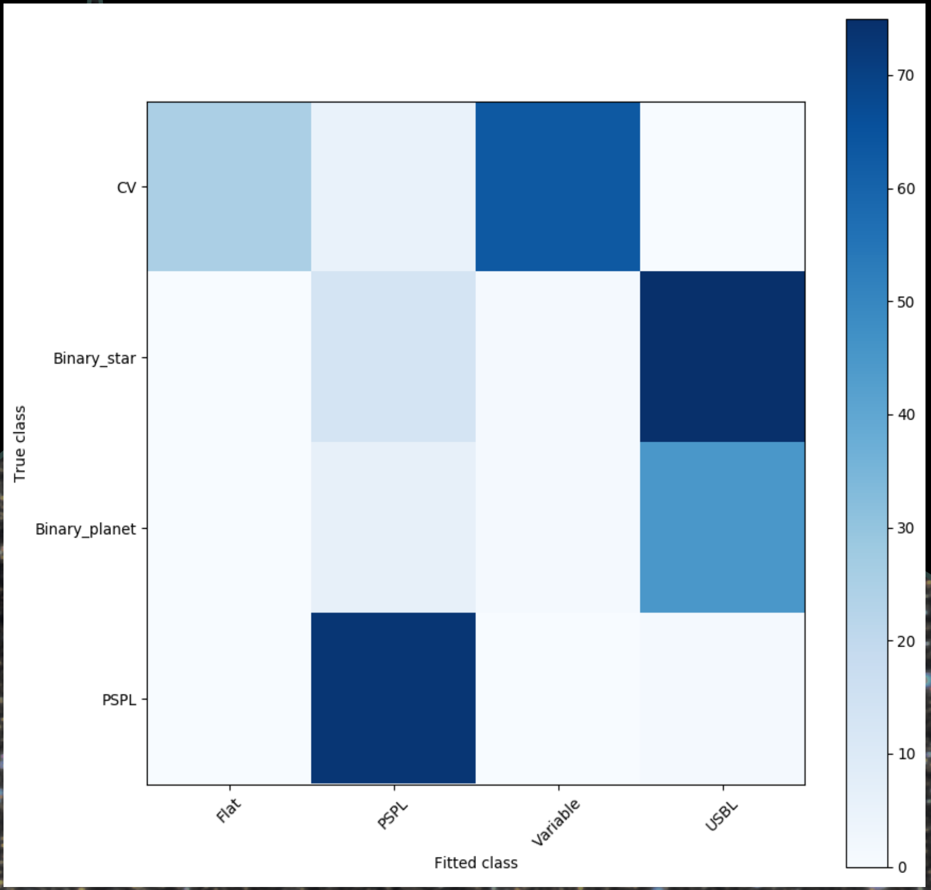
Rachel, Matthew and panel members not permitted to participate in teams

A github organization was provided as a public forum through which questions could be posed and answered.

Evaluating Challenge Entries

Programmatic comparison between each team's entry and simulation master table





Evaluating Challenge Entries

Programmatic comparison between each team's entry and simulation master table

Comparison of fitted parameters for Team1

The table below compares the parameters obtained during the fitting process (black) with the true parameters used to simulate the datasets (grey, italics)

If a team provided several alternative models for a single dataset, these are represented by multiple entries with the same ModelID

'None' entries represent values missing from the team's table entry data.

Back to entry summary

ModelID	Class	t0	tE	u0	rho	piE	fs_W	fb_W	fs_Z	fb_Z	s
ulwdc1_002	PSPL *Binary_star*	2459650.03355 ± 0.2432288499 *2459648.31840479*	55.2527260873 ± 8.4917345258 *36.8514*	0.7818939997 ± 0.1701036718 *2.16671*	None ± None *0.00104122*	None ± None *0.0729903*	59.2011595983 ± 23.7946204763 *0.413613*	1437.78721782 ± 23.776920539 *2.852503507111193*	45.529139359 ± 18.3081887461 *0.557115*	810.515447624 ± 18.29614764 *1.558477832946977*	None ± Non *0.325677*
ulwdc1_004	USBL *Binary_planet*	2460024.22001 ± 0.0820564796 *2460024.31690896*	10.5138334848 ± 0.3966804041 *6.6877*	0.7376195568 ± 0.049117409 *-1.52343*	0.0293813664 ± 0.0061434232 *0.00120717*	None ± None *0.171343*	47.4894608478 ± 5.2688112877 *0.607462*	304.554801836 ± 5.2679127284 *1.2773515524873547*	15.9822566229 ± 1.7903792031 *0.744427*	79.7004473245 ± 1.7911315991 *0.8446015179946584*	1.843624606 ± 0.032415774 *2.48124*
ulwdc1_005	PSPL *PSPL*	2459816.21546 ± 0.0016982558 *2459816.21264422*	1.7893880493 ± 0.0968700982 *1.85542*	0.074382326 ± 0.0051480273 *-0.0712711*	None ± None *0.00280589*	None ± None *0.321621*	12.1824876967 ± 0.8667417313 *0.0206938*	551.305006801 ± 0.8649149434 *1188.2484362550706*	4.3035231358 ± 0.3155290846 *0.0242225*	158.888516415 ± 0.3169077042 *866.6305304269218*	None ± Non *None*
ulwdc1_006	USBL *Binary_star*	2459651.85513 ± 0.00014 *2459651.8548187*	5.7205166855 ± 0.1265098915 *5.31401*	0.0052095829 ± 0.000121 *-0.00588459*	0.000314 ± 0.00045 *0.000772343*	None ± None *0.0902075*	4.5936183648 ± 0.1093460461 *0.0684179*	69.3814569304 ± 0.1078596445 *106.46494240328049*	0.707314727 ± 0.0180179795 *0.102309*	6.7651987964 ± 0.0191512543 *47.72488953238265*	0.101764207 ± 0.002037278 *0.113947*
ulwdc1_008	USBL *Binary_planet*	2458728.97888 ± 0.0782798178 *2458729.060358196*	24.6969140599 ± 0.012271923 *26.2846*	0.9553988635 ± 0.0062206637 *0.891991*	0.0060379147 ± 0.005790243 *0.000547291*	None ± None *0.0852794*	269.887440907 ± 3.2659930145 *0.117726*	1722.52821132 ± 3.2748087289 *37.82059169457552*	130.963101858 ± 1.8441740261 *0.123387*	798.231271522 ± 1.8660343201 *34.64534799491492*	1.964981072 ± 0.005348349 *1.92159*
ulwdc1_009	USBL *Binary_star*	2460010.3793 ± 0.0111140771 *2459971.19486356*	68.5770508011 ± 1.0854118002 *74.0461*	0.1339243004 ± 0.0036791371 *-0.455631*	0.0002158844 ± 19562.5973645534 *0.00113586*	0.294186215297 ± 0.0413956007609 *0.0837776*	315.271470443 ± 8.5666518598 *0.306943*	900.743222029 ± 8.5658655513 *4.127217676764982*	163.280300414 ± 4.4375477854 *0.280945*	524.529954616 ± 4.4383429874 *5.0269384536568955*	0.356605725 ± 0.002179945 *4.65476*
ulwdc1_012	USBL *Binary_planet*	2458598.60756 ± 0.00207	16.1039968738 ± 0.0218605962	0.0475759283 ± 7.59e-05	0.00166 ± 0.000588	None ± None *0.0714066*	570.95653707 ± 1.3587459961	620.137900987 ± 1.3535267983	336.753036923 ± 0.840082427	308.858865384 ± 0.8431768844	1.035444269 ± 0.001636715

Evaluating Challenge Entries

Comparison of simulated/true parameters highlights weaknesses (some known) in modeling process, e.g. tendency for u_0(fitted) > u_0(true)

Comparison o

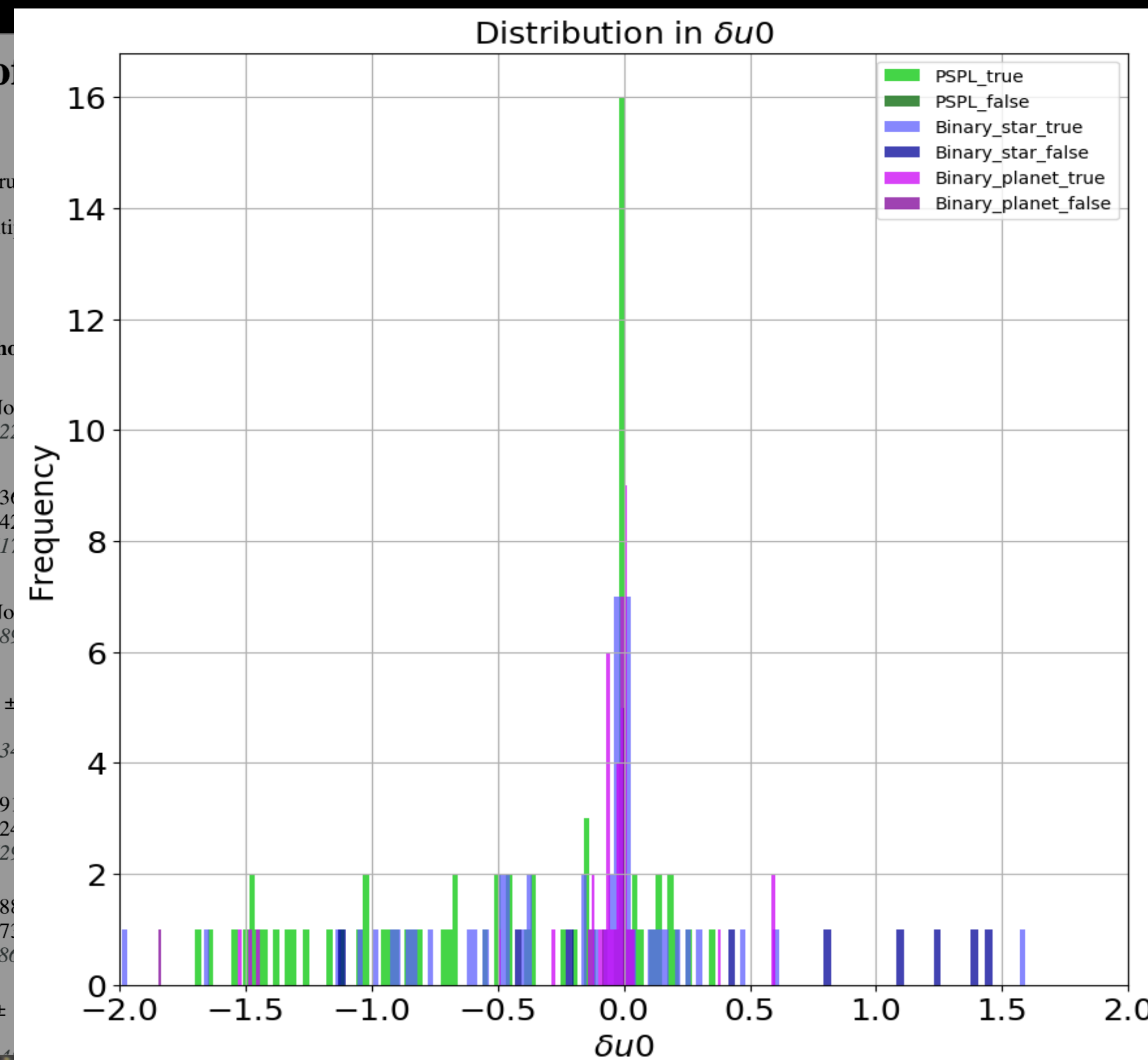
The table below compares the parameters obtained during the fitting process (black) with the tru

If a team provided several alternative models for a single dataset, these are represented by multi

'None' entries represent values missing from the team's table entry data.

Back to entry summary

ModelID	Class	t0	tE	u0	rh
ulwdc1_002	PSPL *Binary_star*	2459650.03355 ± 0.2432288499 *2459648.31840479*	55.2527260873 ± 8.4917345258 *36.8514*	0.7818939997 ± 0.1701036718 *2.16671*	None ± No *0.00104122*
ulwdc1_004	USBL *Binary_planet*	2460024.22001 ± 0.0820564796 *2460024.31690896*	10.5138334848 ± 0.3966804041 *6.6877*	0.7376195568 ± 0.049117409 *-1.52343*	0.0293813 0.0061434 *0.0012071*
ulwdc1_005	PSPL *PSPL*	2459816.21546 ± 0.0016982558 *2459816.21264422*	1.7893880493 ± 0.0968700982 *1.85542*	0.074382326 ± 0.0051480273 *-0.0712711*	None ± No *0.0028058*
ulwdc1_006	USBL *Binary_star*	2459651.85513 ± 0.00014 *2459651.8548187*	5.7205166855 ± 0.1265098915 *5.31401*	0.0052095829 ± 0.000121 *-0.00588459*	0.000314 ± 0.00045 *0.0007723*
ulwdc1_008	USBL *Binary_planet*	2458728.97888 ± 0.0782798178 *2458729.060358196*	24.6969140599 ± 0.012271923 *26.2846*	0.9553988635 ± 0.0062206637 *0.891991*	0.0060379 0.0057902 *0.0005472*
ulwdc1_009	USBL *Binary_star*	2460010.3793 ± 0.0111140771 *2459971.19486356*	68.5770508011 ± 1.0854118002 *74.0461*	0.1339243004 ± 0.0036791371 *-0.455631*	0.0002158 19562.597 *0.0011358*
ulwdc1_012	USBL *Binary_planet*	2458598.60756 ± 0.00207	16.1039968738 ± 0.0218605962	0.0475759283 ± 7.59e-05	0.00166 ± 0.000588



Evaluating Challenge Entries

Panel members awarded grades out of 5 in each category

- Accuracy of fitted model parameters
- Software/modeling process efficiency/scalability
- Innovations in approach
- Broadening the field

Each team will receive written feedback regarding the panel's conclusions

Some important but hard-to-evaluate criteria

True benchmarking not implemented for logistical reasons

Panel relied on documentation to evaluate process and innovative aspects

Evaluation supplemented by questionnaire to all teams, requesting specific information regarding e.g. computing resources used

Team Credits

Team 1	Contact: Etienne Bachelet	Markus Hundertmark Daniel Godines Charlotte Fling
Team 2	Contact: Valerio Bozza	
Team 3	Vandylions Contact: Geoffrey Bryden	Savannah Jacklin Rob Siverd Keivan Stassun Ryan Oelkers
Team 4	Contact: Clément Ranc	Arnaud Cassan Richard K. Barry Esther Euteneuer Stela Ishitani Silva Yiannis Tsapras

Data Challenge Results

Accuracy in fitted parameters

Team	Combined scores	Rank
Team 1	16.17	1
Team 2	14.5	2
Team 3	7.83	4
Team 4	11.0	3

Std.dev 3.72

Overall, when events were properly classified, model parameters could be accurately derived, noting known weaknesses.
Future work to investigate "un-modelable" events
Classification problem non-trivial, particularly for subtle anomalies

Data Challenge Results

Software/modeling process efficiency

Team	Combined scores	Rank
Team 1	13.5	1
Team 2	11.5	2
Team 3	9.5	4
Team 4	11.0	3

Std.dev 1.65

All teams used publicly available software
New approaches to classification/detection in development, but early stage
Effective progress on question of scalability, but room for improvement
At least two teams required laptops/workstations rather than cluster computers

Data Challenge Results

Innovation

Team	Combined scores	Rank
Team 1	14.5	3
Team 2	15	2
Team 3	8	4
Team 4	17	1

Std.dev 3.90

Significant effort invested into development of modern, open-source software
Some welcome attempts made to trial non-standard techniques
Evaluation dependent on documentation provided

Data Challenge Results

Broadening the field

Team	Combined scores	Rank
Team 1	12	3
Team 2	4.5	4
Team 3	14.5	1
Team 4	13.0	2

Std.dev 4.45

All but one of the teams included students and/or researchers whose previous work is primarily outside microlensing
All teams included established microlensers
More work to do to bring in “fully” new teams

Challenge 1 Next Steps

Panel written feedback will be sent to each team today

Paper documenting the challenge and entries underway

- All teams invited to collaborate
- Teams encouraged to send feedback on difficulties encountered
- Further analysis underway of team results, cause of “pathological” cases etc.

All teams are encouraged to publish independent papers as well

Github organization will remain open

- Challenge 1 dataset available for future tests etc
- Forum for questions

Lessons learned

While the processing of large datasets will be a concern for WFIRST, meaningful comparison between teams is difficult without formal benchmarking

- Requires standardized computing facilities
- Could be done with a cloud-based server and virtual environments but some cost associated with this.
- Best achieved in a hackathon-style targeted “mini-challenge” workshop

Lessons learned

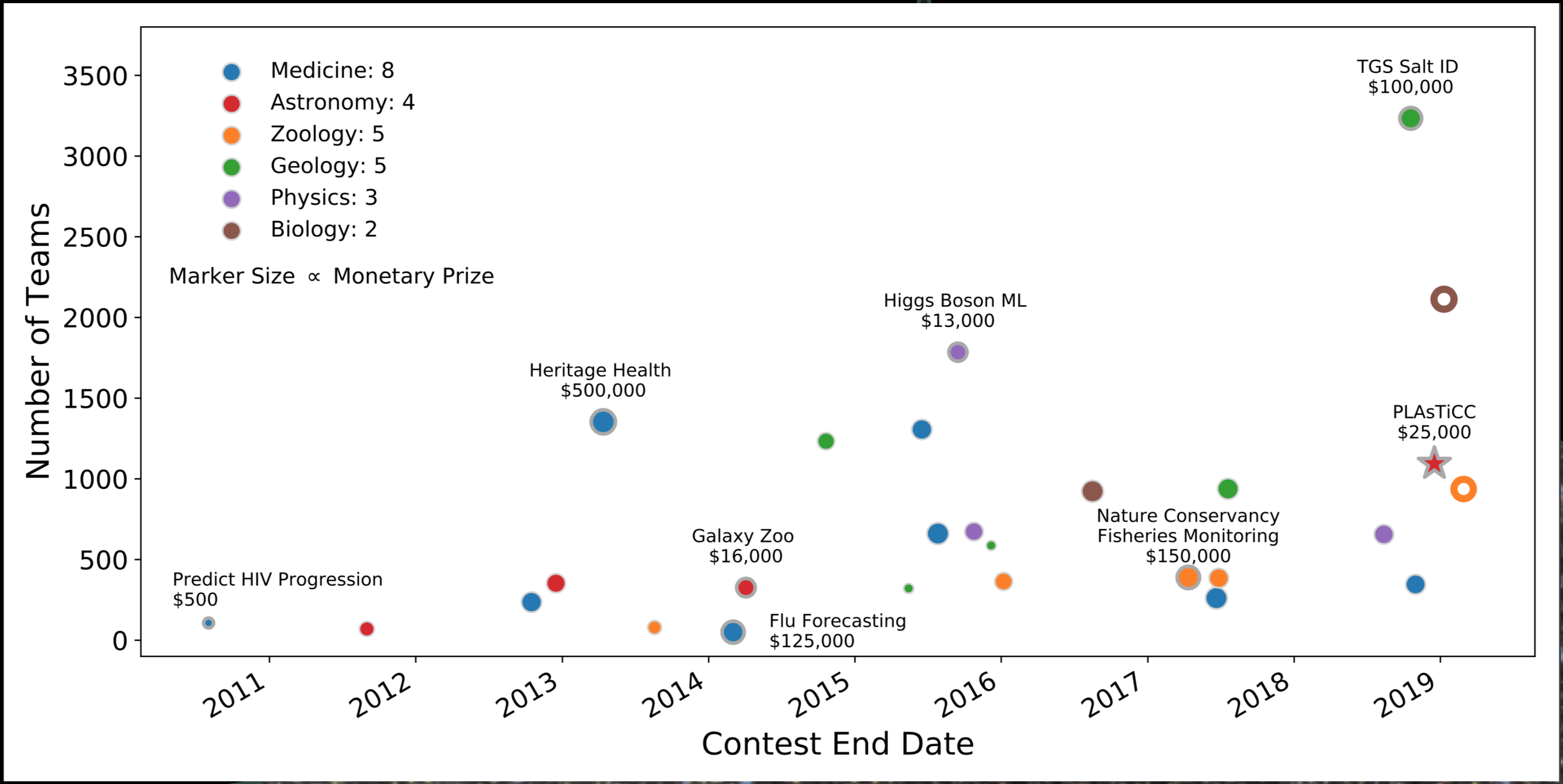
Attracting researchers from outside astronomy/exoplanets was difficult, despite publicizing the challenge on a number of astro-statistics forums

- Recent LSST data challenge used Kaggle platform
  - attracted 1094 teams, most non-astro
- Drawbacks:
  - cost – prizes expected
  - high overhead to prepare challenge dataset to meet platform requirements, avoid "leakage"
  - Kaggle is really designed for supervised classification challenges

Lessons learned

Attracting researchers from outside astronomy/exoplanets was difficult, despite publicizing the challenge on a number of astro-statistics forums

Entrants in recent Kaggle-run Data Challenges



Credit: Gautham Narayan

Next Challenge(s)

Photometry Challenge

WFIRST MicroSIT in the process of planning/developing data handling tools
Timely to address optimal photometry from WFIRST simulated images

Data simulation work in progress; announcement will be made

Next Challenge(s)

Photometry Challenge

WFIRST MicroSIT in the process of planning/developing data handling tools
Timely to address optimal photometry from WFIRST simulated images

Data simulation work in progress; announcement will be made

Mini-challenges

Short timescale (~week) challenges targeting a specific issue
Options for resources such as cloud-servers
Coordinated with other meetings as hack sessions

Thank you, participants and panel

Tiffany Meshkat and IPAC for support of the challenge and public codes